How to use this book

R. Heller · Date: 2026-07-29 · Last changed: 2026-08-31

Standing methodological rules

Three rules apply throughout and are not restated in each chapter.

1. **Every stochastic result is reported across at least 20 seeds, and across more where the spread requires it.** t-SNE and UMAP are seed-dependent; a single run is an anecdote. Twenty is a floor, not a design: the count each method needs is measured, by inverting the 0.02 reportable difference against that method's within-cell spread, and a method needing more gets more. `scripts/measure-seed-budget.R` does the measuring and `data/processed/seed-budget.rds` is its record. Seeds are fixed in `_common.R` as `BENCH_SEEDS` and disclosed with every figure.

   Where an artefact is produced below the floor — a pilot, or an arm whose cost forbids it — the count is named in the text that reports it, and `scripts/check-artefact-producers.R` will not let it pass unnamed.

2. **Every θ-dependent result is reported as a curve, never as a single number.** Fold angle is a continuous difficulty parameter. Reporting one value invites cherry-picking, and hides exactly the regime changes the benchmark exists to expose.

3. **Every figure carries `fig.alt`.** No result is encoded by colour alone.

What you need

R 4.1 or newer and the packages pinned in `renv.lock`. Run `renv::restore()` once after cloning. There is nothing else to install: the pattern constructors, the sampler and the evaluation metrics are plain scripts in `R/`, sourced by `_common.R` when the book renders. The code that produces every result in this book is in the same repository as the book, at the same commit.

The Chapter 10 benchmark grid takes hours to compute. It is precomputed and committed under `data/processed/`; `scripts/run-benchmark-grid.R` is kept as the provenance record rather than run during the build.

How the chapters are shaped

Every body chapter uses the same nine sections, in the same order, with the same anchors. The anchors are namespaced by a short chapter mnemonic — `geom-`, `intro-`, `eval-` — because pandoc slugs headings by their text, and twelve chapters each carrying a heading called "Results" would otherwise collide and silently renumber each other's cross-references.

```
## What this chapter answers  {#<mn>-question}
## Setup                      {#<mn>-setup}
## Background                 {#<mn>-background}
## <free-wording core>         {#<mn>-core}
## Results                    {#<mn>-results}
## Diagnostics                {#<mn>-diagnostics}
## Where it fails             {#<mn>-limits}
## Reproduce this             {#<mn>-reproduce}
## Further reading            {#<mn>-reading}
```

Only the core section is freely worded — it is the one that carries the chapter's actual argument, and calling it "Core" in every chapter would be worse than useless. Everything else is fixed, and `scripts/lint-chapters.R` fails the build on a chapter that is missing a slot, has them out of order, or anchors one to the wrong namespace.

Read it as a contract with you rather than a template for us. *Where it fails* is in every chapter because every method in this book fails somewhere, and a chapter that could not fill that section would not be finished. *Reproduce this* names the artefact and the script that produced it, so any number in the chapter can be traced to the run that generated it.

Callouts

Five callout classes are defined in `style/style.css`, each with a LaTeX fallback so they survive into the PDF.

- `boxinfo` — context a reader can skip without losing the thread.
- `boximportant` — a result that is easy to misread, or a default that will bite.
- `boxquestion` — something to try, with the answer available in the sources.
- `boxreport` — an aside on provenance, attribution or the state of the literature.
- `boxempty` — an unadorned block, no icon and no colour. It carries the suggested citation in Citing this book.

Write them as knitr block chunks:

```
<div class="boximportant">
<p>umap::umap() defaults to preserve.seed = TRUE.</p>
</div>
```

Never as a raw `<div class="boxinfo">` and never as a pandoc fenced div (`::: boxinfo`). Those two forms are not what the stylesheet targets: they render as unstyled markup in HTML and disappear entirely in PDF, where the boxes are LaTeX environments declared in `style/preamble.tex`. That is how a sibling volume shipped eighty unstyled boxes.

How to read it

Part I builds the object. Part II tests four families of method against it. Part III is the benchmark proper and contains the book's main contribution — Chapter 9, where the evaluation metrics themselves are checked against exact truth. Part IV asks whether any of it changes a real decision.

Readers who only want the benchmark can start at Chapter 8 and refer back to Chapter 3 for the generative model.