Software overview: SUMMER, surveyPrev, sae4health, and more

Zehang Richard Li
Department of Statistics
University of California, Santa Cruz

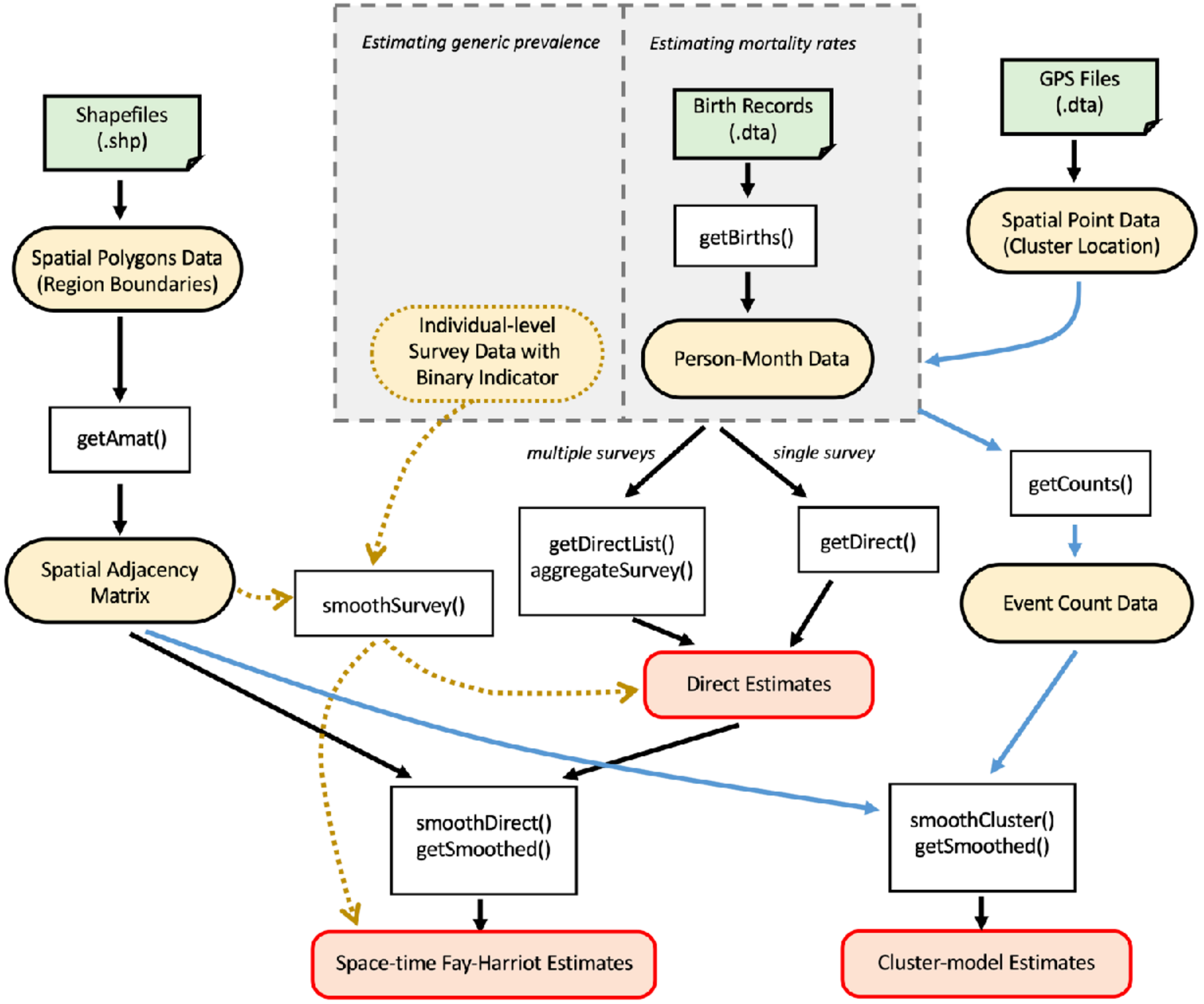
STAB Weekly Meeting
10/7/2025

History

- 2017: SUMMER (Small-area Under-five Mortality Model Estimation in R)
  - https://cran.r-project.org/web/packages/SUMMER/index.html
  - Initially reproducing Laina et al (2015). SAE for child mortality estimation using space-time Fay-Herriot models.
  - Over time, more functionalities are added:
    - Cluster-level models
    - Generic SAE for binary outcome
    - More flexible space-time components
    - Benchmarking
    - …

SUMMER

- Current: SUMMER (Small-Area-Estimation Unit/Area Models and Methods for Estimation in R)
- Version 2.0.0 on CRAN (Jan 2025). Some minor updates on GitHub.
- Primarily useful for space-time child mortality models.
- Space-time or space-only model for generic binary indicator: *smoothSurvey().*



SUMMER

- Two other functions in SUMMER are imported into the survey package by Peter Gao.
  - *smoothArea()*
  - *smoothUnit()*
- They follow survey package workflow more closely, i.e., specifying survey design and then build model on it.

survey: Analysis of Complex Survey Samples

Summary statistics, two-sample tests, rank tests, generalised linear models, cumulative link models, Cox models, loglinear models, an
Variances by Taylor series linearisation or replicate weights. Post-stratification, calibration, and raking. Two-phase subsampling desig

Version:	4.4-8
Depends:	R (≥ 4.1.0), grid, methods, Matrix, survival
Imports:	stats, graphics, splines, lattice, minqa, numDeriv, mitools (≥ 2.4), Rcpp (≥ 0.12.8)
LinkingTo:	Rcpp, RcppArmadillo
Suggests:	foreign, MASS, KernSmooth, hexbin, RSQLite, quantreg, parallel, CompQuadForm, DBI, AER, SUMMER (≥ 1.
Published:	2025-08-28
DOI:	10.32614/CRAN.package.survey
Author:	Thomas Lumley [aut], Peter Gao [aut], Ben Schneider [aut], "Thomas Lumley" [cre]
Maintainer:	"Thomas Lumley" <t.lumley at auckland.ac.nz>
License:	GPL-2 | GPL-3
URL:	http://r-survey.r-forge.r-project.org/survey/
NeedsCompilation:	yes
Citation:	survey citation info
Materials:	NEWS
In views:	OfficialStatistics, Survival
CRAN checks:	survey results

Documentation:

Reference manual:	survey.html , survey.pdf
Vignettes:	Small area estimation (source) Estimates in subpopulations (source, R code) Two-phase designs in epidemiology (source, R code) Analysing PPS designs (source, R code) Pre-calibrated weights (source, R code) Quantile rules (source) A survey analysis example (source, R code)

SUMMER: Small-Area-Estimation Unit/Area Models and Methods for Estimation in R

Provides methods for spatial and spatio-temporal smoothing of demographic and health indicators using survey data, with p
al. (2019) <doi:10.1371/journal.pone.0210645>, Wu et al. (DHS Spatial Analysis Reports No. 21, 2021), and Li et al. (2023

Version:	2.0.0
Depends:	R (≥ 3.5)
Imports:	survey, stats, spdep, survival, ggplot2, scales, utils, Matrix, reshape2, viridis, sp, sf, shadowtext, g
Suggests:	INLA, sn, knitr, rmarkdown, readstata13, patchwork, rdhs, R.rsp, sae, dplyr, tidyr, raster, fmesher
Published:	2025-01-07
DOI:	10.32614/CRAN.package.SUMMER
Author:	Zehang R Li [cre, aut], Bryan D Martin [aut], Yuan Hsiao [aut], Jessica Godwin [aut], John Paige
Maintainer:	Zehang R Li <lizehang at gmail.com>
BugReports:	https://github.com/richardli/SUMMER/issues
License:	GPL-2 | GPL-3 [expanded from: GPL (≥ 2)]
URL:	https://github.com/richardli/SUMMER, https://richardli.github.io/SUMMER/
NeedsCompilation:	no
Additional_repositories:	https://inla.r-inla-download.org/R/testing/
Materials:	README, NEWS
In views:	OfficialStatistics
CRAN checks:	SUMMER results

Documentation:

Reference manual:	SUMMER.html , SUMMER.pdf
Vignettes:	Specifying cluster-level model for mortality estimation (source) Generic small area estimation (source) Case Study: Estimating Subnational U5MR using DHS data (source) Estimating Subnational U5MR using Simulated Data (source)

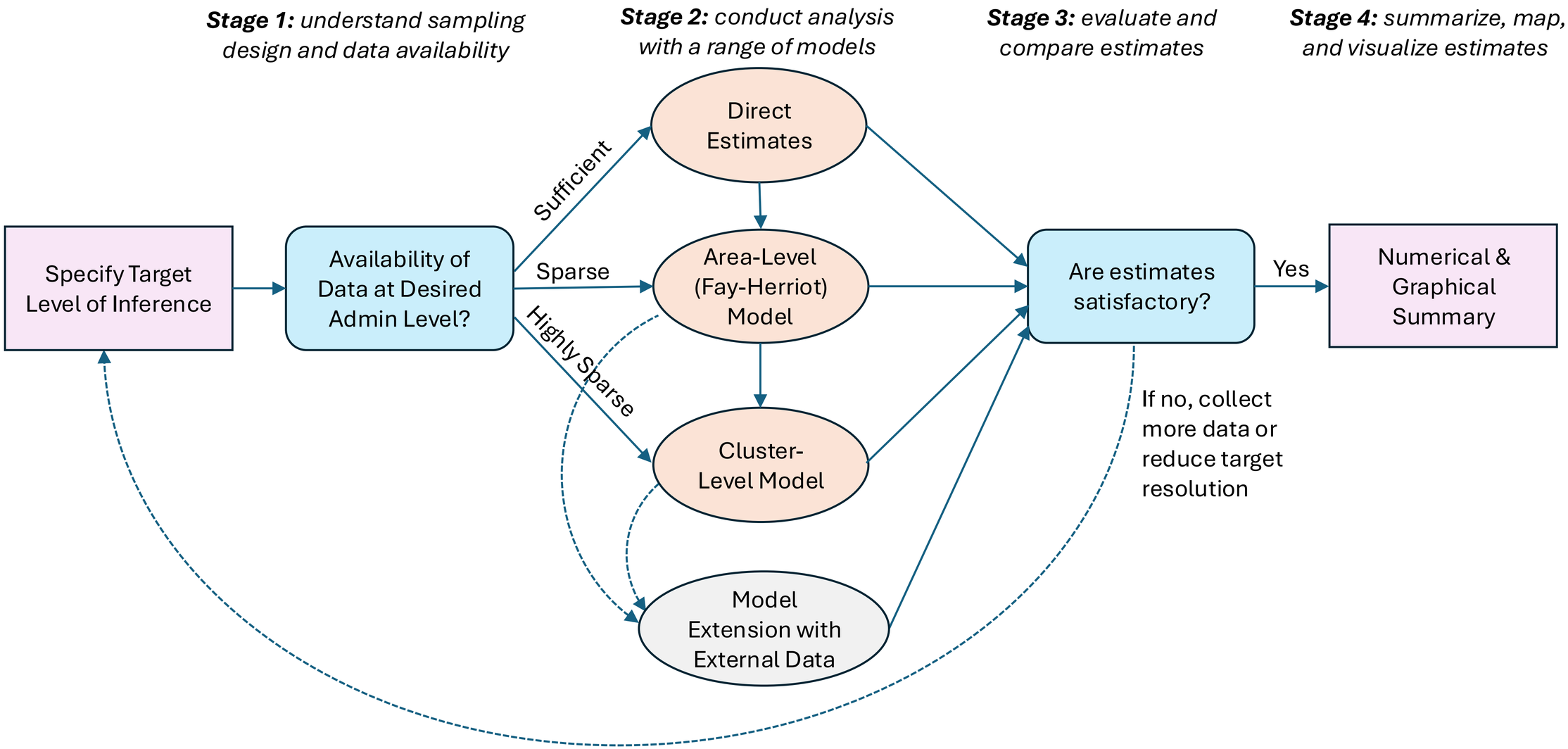
SUMMER: some notes

- There are many small utility functions we have added in SUMMER over time, e.g.,
  - Options to handle Ethiopian calendar being 92 months behind the Gregorian (western) calendar.
  - Options to handle older DHS where deaths were only recorded in year rather than month.
- There are some useful mapping / visualization functions.
- There is a collection of functions for simulating data from population grid, developed by Johnny Paige.

surveyPrev

- With more and more demand for modeling generic indicator (there are many of them!) It turns out mapping the DHS recodes into binary indicators is a huge pain.
- We started the idea of a new package surveyPrev in 2023. Qianyu Dong has led the development with contributions from Yunhan Wu, Jieyi Xu, Andrea Boskovic, etc.
- The idea is to do everything in R:
  - Load data and map cluster to region using spatial polygons.
  - Automatic processing of binary indicators (initially 22, now 160+) & Customized indicator definition.
  - Functions to process external data, e.g., population fractions.
  - Direct estimation, FH, cluster-level model at multiple Admin levels. Automatic aggregation of results.
  - Visualization and reporting.

surveyPrev


Stage 1: *understand sampling design and data availability*
Stage 2: *conduct analysis with a range of models*
Stage 3: *evaluate and compare estimates*
Stage 4: *summarize, map, and visualize estimates*
Specify Target Level of Inference
Availability of Data at Desired Admin Level?
Sufficient
Sparse
Highly Sparse
Direct Estimates
Area-Level (Fay-Herriot) Model
Cluster-Level Model
Model Extension with External Data
Are estimates satisfactory?
Yes
Numerical & Graphical Summary
If no, collect more data or reduce target resolution

surveyPrev: the indicators

- surveyPrev is currently at version 1.0.0 on CRAN (April 2024 version). It's very outdated! Until we update on CRAN, you should always install from GitHub.

- A very related source for more indicators is the DHS indicator repository: https://github.com/DHSProgram/DHS-Indicators-R

  - We have both manually and automatically extracted functions from this repository, but currently still only a small fraction of indicators here.

- More DHS indicators are at: https://dhsprogram.com/data/statcompiler.cfm

- More details of indicators and how they change over time at: https://www.dhsprogram.com/publications/publication-dhsg1-dhs-questionnaires-and-manuals.cfm

surveyPrev & SUMMER: dependency and comparisons

- Internally, surveyPrev uses some functions from SUMMER to fit models. We may decouple some functions in the future as the focuses of the two packages become more different.

- There are more built-in support in surveyPrev for aggregation and cross-Admin comparison. Currently it's always two levels, but Admin-1 and Admin-2 specifications can be swapped with, e.g., Admin-1 and Admin-3; Admin-2 and Admin-3, etc.

- There are slightly more support for non-standard stratification in SUMMER. surveyPrev currently requires two levels of stratification, i.e., urban and rural. *This may change in the future!* SUMMER allows more levels (but not extensively tested).

- For child mortality related composite indicators (e.g., U5MR), there is no spatial-only unit-level model implemented in SUMMER. *This may change in the future!*

Where to get help

- SUMMER website: https://richardli.github.io/SUMMER/
- The four vignettes for SUMMER on CRAN
- SAE vignette in survey package
- surveyPrev workflow paper: https://arxiv.org/abs/2504.16435 and the reproducible report that include all data processing steps, e.g., manual check of misclassification, extracting population and urban/rural fractions: https://github.com/dqianyu/WorkflowPaper
- Function source codes is usually a great way to check out what's going on.

Beyond R

- Over time, we have given workshops to teach people how to use our R packages (and initially, the step by step model building using INLA as well!)
- It was clear that R created many more problems! They never install and run without issues…and the majority of participants will either never open R again or immediately hit with code errors when trying it out afterwards.
- It was clear we need software interface without coding from command line.
- Tracy Dong developed an earlier version of shiny app for SUMMER.
- Then in 2024, Yunhan Wu created the greatest shiny app that rules all R packages…

Getting Started
Country Specification
Data Upload
Model Fitting
Result Visualization
Result Tabulation
Report Generation
Tool Kit

Country Meta Data Input

Choose a country *
Burkina Faso

Choose survey year *
2021

Choose an indicator group *
Chapter 10 - Child Health

Choose an indicator *
Children with ARI for whom advice or treatment was sought

Choose Admin levels for analysis *
- [x] National
- [x] Admin-1
- [x] Admin-2
- [x] Admin-3

Check out maps for Admin levels
National

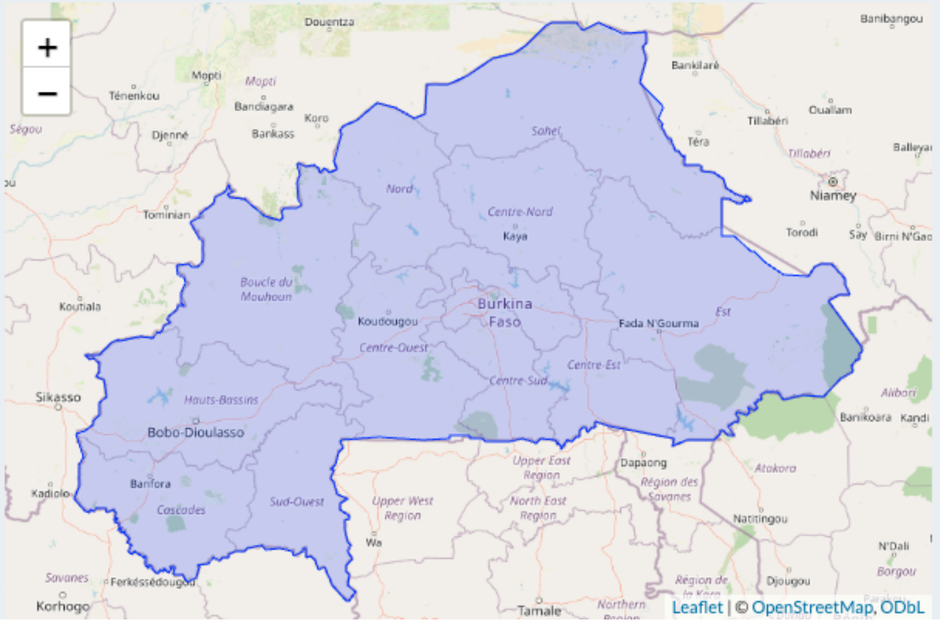
Interactive Map Enabled

You've selected **Burkina Faso** with survey in **2021**, to estimate **Children with ARI for whom advice or treatment was sought** (see detailed definition here).
You intend to conduct analysis at **National, Admin-1, Admin-2 and Admin-3** level(s).

Please review the table and map below for your Admin level selections. Choose a different level to display if necessary.

	National	Admin-1	Admin-2	Admin-3
Number of Regions	1	13	45	351

The map below displays **National** boundaries of **Burkina Faso**.



sae4health package

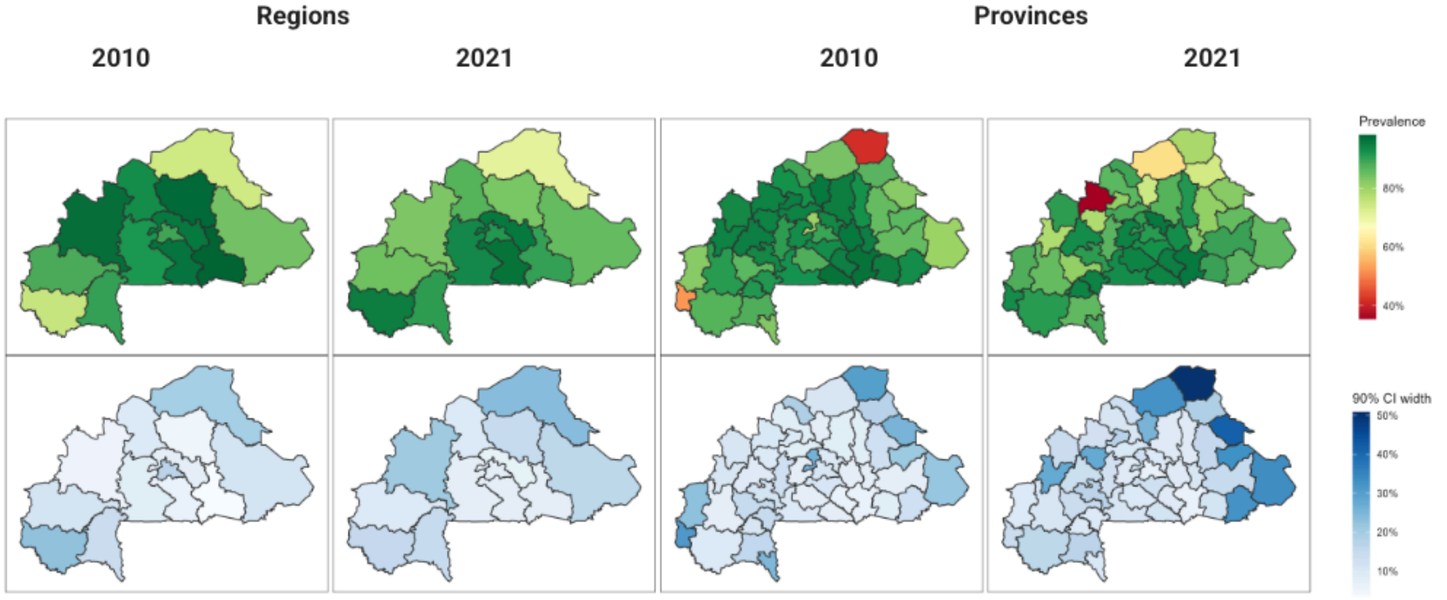
- The shiny app is currently at https://rsc.stat.washington.edu/sae4health/
- It's directly running from the sae4health package.
- You can deploy it locally or use the server version. In the local version, you have to upload the DHS data manually (with instructions). In the server version, all DHS data are read directly from UW server.
- Many functionalities in sae4health package, including comparison with DHS API, descriptive plots, interactive visualizations, covariate upload, report generation, etc.
- Again, the focus is on the right workflow. Users need to pick the models (with some assistance with data sparsity checks) and the results are then computed.
- There are also a few variations of the website for WHO, Nigeria, MICS, etc.

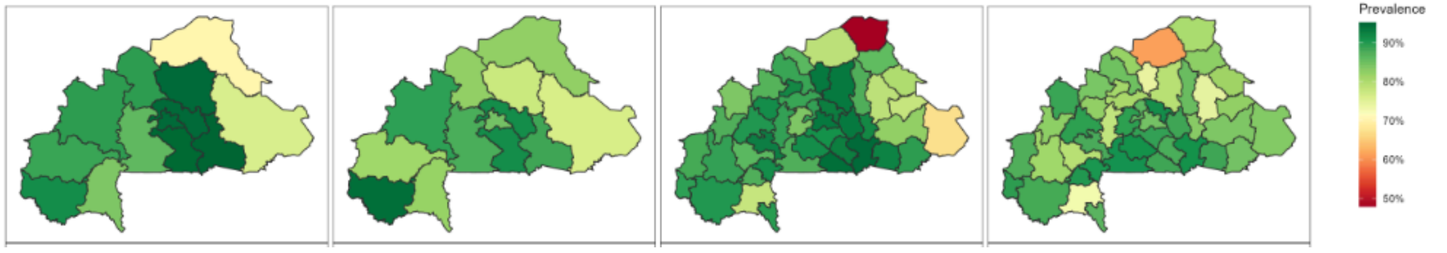
sae4lmic website

- What if modeling one indicator at a time and going through the model choice steps is still too much work?
- Over the summer, we have created yet another website: https://sae4lmic.stat.uw.edu/
- It's based on only pre-computed results for selected indicators and countries, deployed by MkDocs and GitHub action (so .md files + saved plots only).
- We run analysis with multiple models and select ones we trust to put on the website.
- It's still very early and we are learning how to make it efficient to disseminate estimates rather than model tools.

Burkina Faso

Download Report





SUMMERverse —> prevEco?

- R packages
  - SUMMER: processing and modeling child mortality
  - surveyPrev: processing and modeling binary DHS indicators
  - sae4health: interactive workflow
- Websites
  - DHS app, MICS app, Multi-survey app, …
  - Result-only websites (with links to modeling shiny apps)

What else?

- A MICS indicator package?
  - Maybe we should split the indicator definitions from surveyPrev and create an indicator processing-only package for DHS, MICS, MIS, etc.?
- A data package: a large collection of DHS survey information:
  - Population counts/fractions, modeled U/R fractions
  - Census information
  - Country- or survey-specific topics, e.g., special maps, non-standard sampling design, etc.

Specific things that may happen soonish

- Child mortality indicators using the surveyPrev workflow.
- Better S3 summary/visualization/documentation of surveyPrev objects.
- More interactive plots.
- Cross validation and model checking.
- SPDE models.

Contribute to packages

- We need more hand!
- What can be useful for R packages:
  - Bug/typo fix, expanded documentation, more examples, …
  - More plotting functions and customization options.
  - Additional models.
- We will standardize the internal objects of surveyPrev so data structure remains the same in long term. If you have made good graphics or summary functions or new model implementations, we want them in the package!
- **How to contribute:** Fork the GitHub repository, and do a Pull Request. Submit Issues on GitHub if you would like to discuss specific changes.

Contribute to shiny app

- Shiny app workflow:
  - Local testing with sae4health package.
  - Same Pull Request workflow on GitHub.
  - Deploy to the UW posit connect *with the right set of dependencies*.
- We also can use more hands on experimental shiny app prototypes spinning off from the main app, and major extensions, e.g., stratified model.
- We are working on a good workflow for people to contribute to these steps.
- There will be a training event at some point.

Contribute to websites

- Same Pull Request workflow on GitHub.
- We need more hands to:
  - Edit and maintain the STAB and Gates website (messing around with HTML, CSS, text, etc.)
  - Help run models and check models for the Gates website.
  - Write blog posts on data analysis and interpretation.
  - …