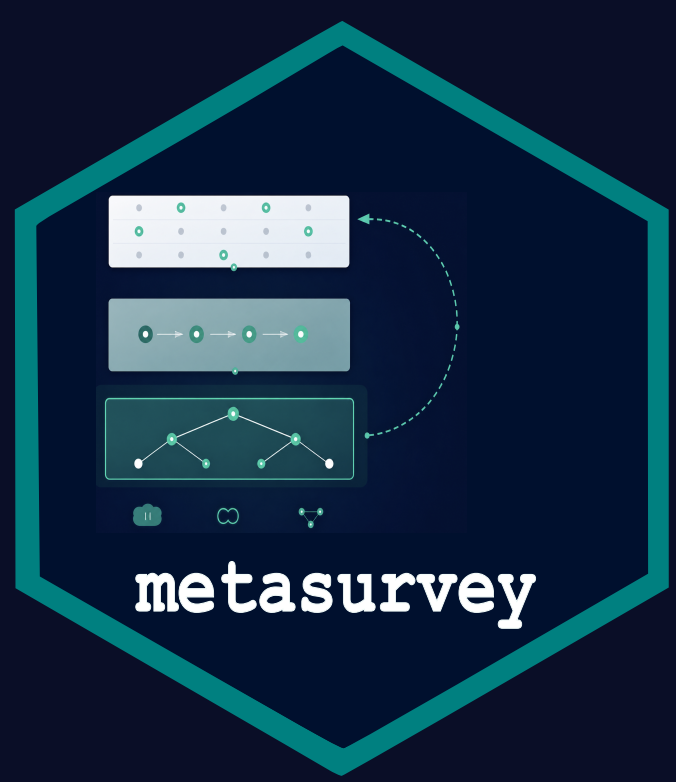




Reproducible Survey Pipelines in R with **metasurvey**

Mauro Loprete, Natalia da Silva, Fabricio Machado

IESTA — Universidad de la República, Uruguay

useR! 2026 · Warsaw, Poland · July 6–9, 2026

useR!
2026

FCEA FACULTAD DE CIENCIAS ECONÓMICAS Y DE ADMINISTRACIÓN

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

► The Pipeline

- National statistical offices produce the **same indicators** from survey microdata — each team writes **ad hoc scripts**, resulting in different code and different results
- **metasurvey** provides a reproducible pipeline:
 - *Steps* — lazy transformations on `data.table`
 - *Recipes* — portable JSON pipelines with dependency graphs
 - *Workflows* — estimation with quality flags
- Three R6 classes support all sampling designs:
 - `Survey` — stratified and cluster designs (`svydesign`)
 - `RotativePanelSurvey` — bootstrap replicate weights
 - `PoolSurvey` — multi-period combined estimates
- Built using only **base R metaprogramming** — no `rlang` dependency

► Get metasurvey

```
install.packages("metasurvey")
```

CRAN	GitHub	Docs	License
v0.0.24	github.com/metasurveyr/ metasurvey	metasurveyr.github.io/ metasurvey	GPL-3

Tech stack: R6, `data.table`, `survey`, `convey`, `Plumber`, `MongoDB`, `Shiny`. 25 bilingual vignettes (EN/ES). CI/CD on Ubuntu, Windows, macOS. Docker deployment.

► Example with metasurvey

```
library(metasurvey)
svy <- survey_empty("ech", "2023") |>
  set_data(microdata) |>
  set_weight(add_weight(annual = "pesoano")) |>
  step_recode(employed,
    POBPCOAC == 2 ~ 1, .default = 0) |>
  step_compute(income_pc = ht11 / ht19) |>
  bake_steps()
```

Steps are **lazy** — recorded but not executed. `bake_steps()` materializes all pending transforms on a `data.table` backend. Uses `substitute()` + `parse()` for expression capture. Supports `compute`, `recode`, `rename`, `remove`, `join`, and `filter` step types.

► Estimation Workflows

`workflow()` wraps `svymean`, `svytotal`, `svyratio`, and `svyby` — returns a tidy `data.table` with **95% CI** and automatic **quality flags** (INE Uruguay standard). Integrated with `convey` for inequality measures: Gini, Theil, FGT, and at-risk-of-poverty.

```
results <- workflow(svy,
  svymean(~employed, na.rm = TRUE),
  svyratio(~po, ~pet),
  svyby(~income_pc, ~gender, svymean),
  convey::svygini(~income_pc)
)
```

indicator	estimate	SE	95% CI	quality
Activity rate	0.632	0.003	[0.626, 0.638]	Excellent
Employment	0.578	0.003	[0.572, 0.584]	Excellent
Unemployment	0.085	0.002	[0.081, 0.089]	Good
Gini (income)	0.387	0.004	[0.379, 0.395]	Excellent

► International Compatibility

Survey-agnostic design tested with household surveys from multiple countries:

Survey	Country	Editions
ECH	Uruguay	40+
EPH	Argentina	10+
CASEN	Chile	5+
PNAD-c	Brazil	3+
CPS	United States	2+
ENIGH	Mexico	2+

► Recipes & Dependency Graphs

Recipes bundle steps + metadata into **shareable JSON** with auto-generated dependency graphs. Each recipe tracks survey type, edition, topic tags, and full variable lineage.

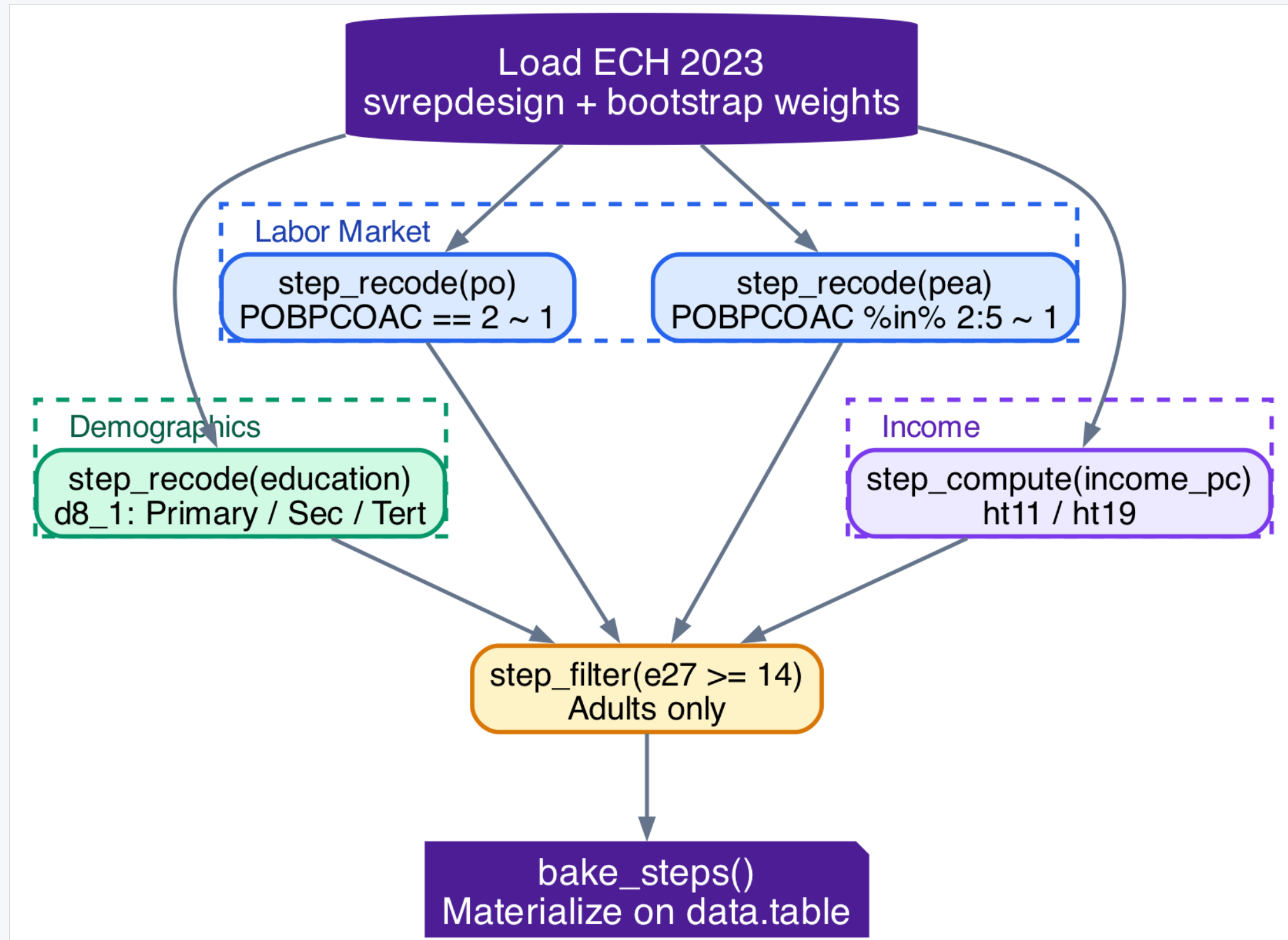


Figure 1: Recipe DAG — ECH 2023: labor, demographics, income

► Stata .do Transpiler

Built-in transpiler converts legacy Stata scripts into native `metasurvey` pipelines. Maps `generate`, `replace`, `recode`, `rename`, `drop`, `keep`, `merge` to step equivalents. We transpiled the **complete historical archive** of Uruguay's ECH:

347
.do files transpiled

34
periods (1981–2022)

100%
coverage

► Shiny Explorer

Interactive web app for **browsing, searching, and executing recipes**. Explore recipes, inspect dependency graphs, run estimation workflows — no R coding required.

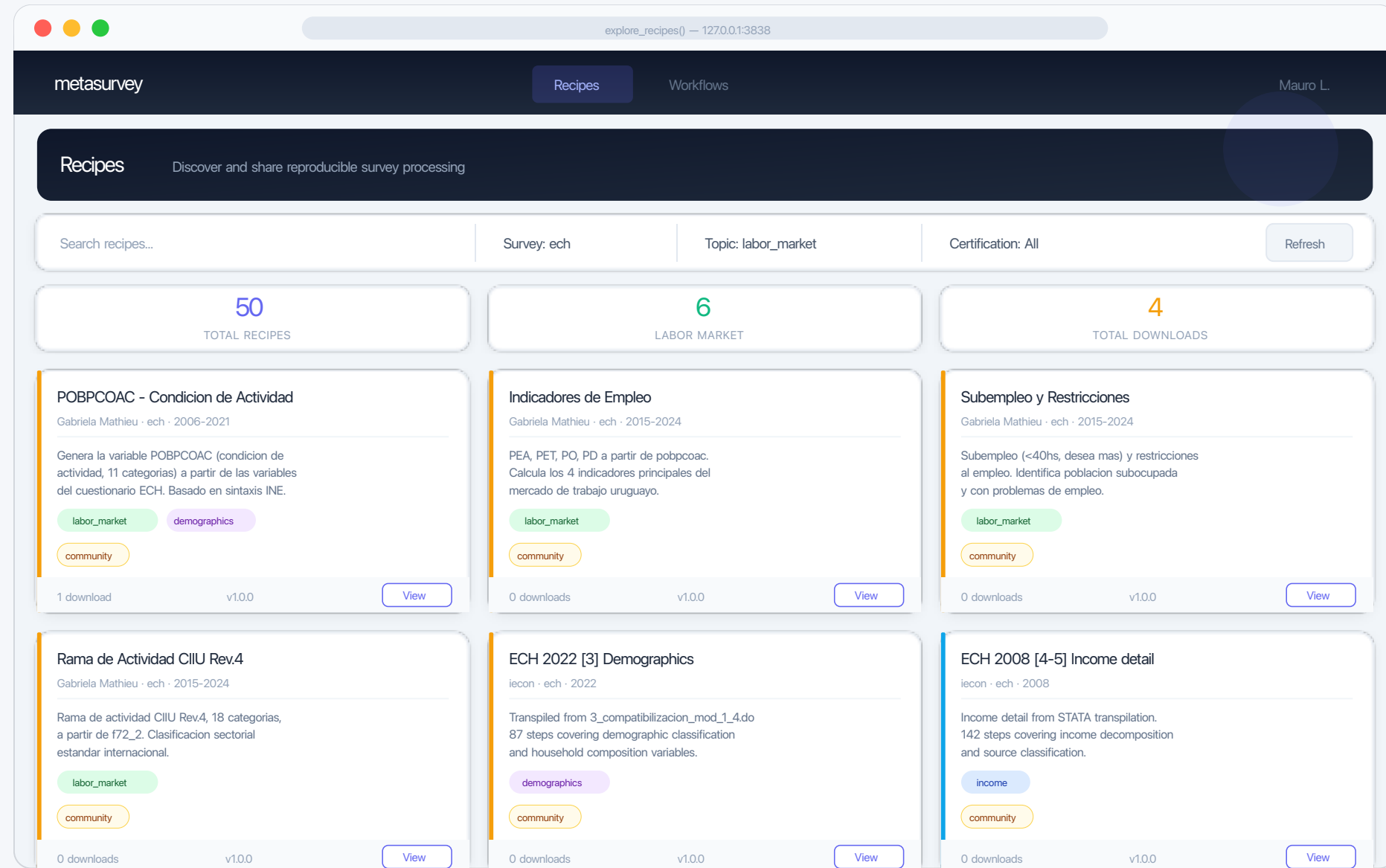


Figure 2: Shiny Explorer — recipe browsing and workflow execution

► Discussion & Future Work

- Steps + Recipes + Workflows reduce **code duplication**: same JSON pipeline, same results across teams and editions
- Stata transpiler proves **real-world complexity**: 40+ years of production code, fully converted
- Future: **rOpenSci** submission, recipe recommendation engine, expanded international coverage

► References



Figure 3: GitHub repository

- T. Lumley. *Complex Surveys: A Guide to Analysis Using R*. Wiley, 2010.
- T. Lumley. *survey: Analysis of Complex Survey Samples*. R package, 2023.
- W. Chang. *R6: Encapsulated Classes with Reference Semantics*. R package, 2021.
- M. Dowle, A. Srinivasan. *data.table: Extension of data.frame*. R package, 2023.