The Journal of Open Source Software

hudlink: Automated ACS–HUD data linking for housing-economics research and analysis

Shane Dabney [1]

1 Florida State University

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 02 August 2025
Published: unpublished

Summary

hudlink is an open-source Python package that compresses a weeks-long tangle of data-wrangling steps into a single command for U.S. housing economists and policy researchers. It

- ingests ACS micro-data via the IPUMS API or local files;
- deterministically imputes missing county codes while preserving survey weights;
- links HUD area-median-income limits (30%, 50%, 80%);
- flags program eligibility and protected-class characteristics;
- merges administrative records for Housing Choice Vouchers, LIHTC, Public Housing, and other programs; and
- exports both CHAS-style county summaries and fully flagged household-level micro-data.

Because public HUD releases supply only pre-aggregated tables, analysts cannot test household-level relationships or audit subsidy allocation for fairness without building bespoke pipelines. **hudlink** removes that barrier, delivering reproducible, analysis-ready data for any U.S. state and any 5-year ACS release.

Statement of need

Housing economists and policy analysts routinely need fine-grained evidence on housing affordability and the reach of federal subsidy programs. Building such datasets from scratch requires locating multiple sources, harmonizing inconsistencies, accounting for missing geography, rescaling survey weights, and coding eligibility rules. For research teams implementing these steps from scratch, the process is time-consuming, error-prone, and difficult to reproduce. **hudlink** automates this workflow. It outputs household-level ACS micro-data already merged with county-specific HUD income thresholds and eligibility flags, plus optional protected-class indicators. A simple, editable configuration lets users re-run the pipeline for new years, states, variables, or HUD programs without changing code.

Implementation

Data sources. IPUMS USA ACS PUMS micro-data (Ruggles et al., 2025) retrieved on demand through the IPUMS Extract API (*IPUMS Microdata Extract API*, n.d.); HUD Area-Median-Income limits for 2009 – 2023 (U.S. Department of Housing and Urban Development, n.d.b);

HUD Picture of Subsidized Households micro-records (U.S. Department of Housing and Urban Development, n.d.a); and the Missouri Census Data Center Geocorr 2012 and 2022 PUMA-to-county crosswalks (Missouri Census Data Center, n.d.a, n.d.b). All non-IPUMS inputs are pre-processed and fetched automatically on the first run.

Pipeline design. Separate modules handle validation, geography harmonization, income cleaning, and eligibility determination. When county IDs are missing, a crosswalk deterministically assigns counties based on PUMA shares, producing weighted copies for split PUMAs and preserving sample design.

Household splitting. An optional procedure (adapted from (Dabney, 2024)) separates multi-family households into constituent family units when overcrowding suggests multiple subsidy-eligible families share the same dwelling. When enabled, hudlink adjusts survey weights to reflect the split families while also preserving the original weights for analyses requiring Census-consistent totals.

Research applications

- **Current Application.** hudlink generalizes and extends the methodology from an earlier pilot, **HCVGAPS** (Dabney, 2024). Despite its more limited scope, that pilot script was adopted and used in published research to audit federal housing-program efficacy and estimate projected federal costs of implementing certain policy recommendations, (Taylor et al., 2024). Building on this momentum, `hudlink` has been reengineered to support all U.S. states and territories, any ACS 5-year release, and all major HUD programs, enabling even more robust nationwide, longitudinal, and cross-program analyses.
- **Algorithmic-bias audits.** Analysts can use hudlink to construct a program-eligibility pool with variables such as race, ethnicity, disability status, veteran status, and education, merge this with HUD recipient micro-data, and then compare the resulting distributions or run simulated draws to test for statistically significant asymmetries.
- **Housing Policy and Planning Research.** Using hudlink, researchers or state housing agencies can quickly estimate unmet housing assistance need by county, identify geographic disparities in program coverage, or evaluate policy changes over time.

Acknowledgements

This project was supported in part by the Institute for Humane Studies (Grant No. IHS018262). I am grateful for the outstanding research assistance of FSU UROP students Iris Bui and Mira Scannapieco, as well as the valuable feedback and support from Parker Ridaught, Eliza Terziev, and Max Blumenfeld. I also thank the DeVoe L. Moore Center at Florida State University for providing resources and facilities throughout this project's development. # References

Dabney, S. (2024). Calculating county-level housing choice voucher gaps: A methodology. *Cityscape*, *26*(2), 401–412.

IPUMS microdata extract API. (n.d.). https://developer.ipums.org/.

Missouri Census Data Center. (n.d.a). *Geocorr 2018: Geographic correspondence engine*. https://mcdc.missouri.edu/applications/geocorr2018.html.

Missouri Census Data Center. (n.d.b). *Geocorr 2022: Geographic correspondence engine*. https://mcdc.missouri.edu/applications/geocorr2022.html.

Ruggles, S., Flood, S., Sobek, M., Backman, D., Chen, A., Cooper, G., Richards, S., Rodgers, R., Schroeder, J., & Williams, K. C. W. (2025). *IPUMS USA: Version 16.0 [dataset]* [Data set]. IPUMS. https://doi.org/10.18128/D010.V16.0

Taylor, C., Staley, S., Dabney, S., Miller, E., Terziev, E., Silva, G., & Nair, J. (2024). *Unlocking*

potential by elevating housing vouchers: Best practices and policy insights. Florida Policy Project.

U.S. Department of Housing and Urban Development. (n.d.a). *A picture of subsidized households: General description of the data and bibliography*. https://www.huduser.gov/portal/datasets/assthsg/statedata98/descript.html.

U.S. Department of Housing and Urban Development. (n.d.b). *Income limits*. https://www.huduser.gov/portal/datasets/il.html#year2022.