R/Medicine 2025

An R Consortium Conference

★ June 9-13, 2025
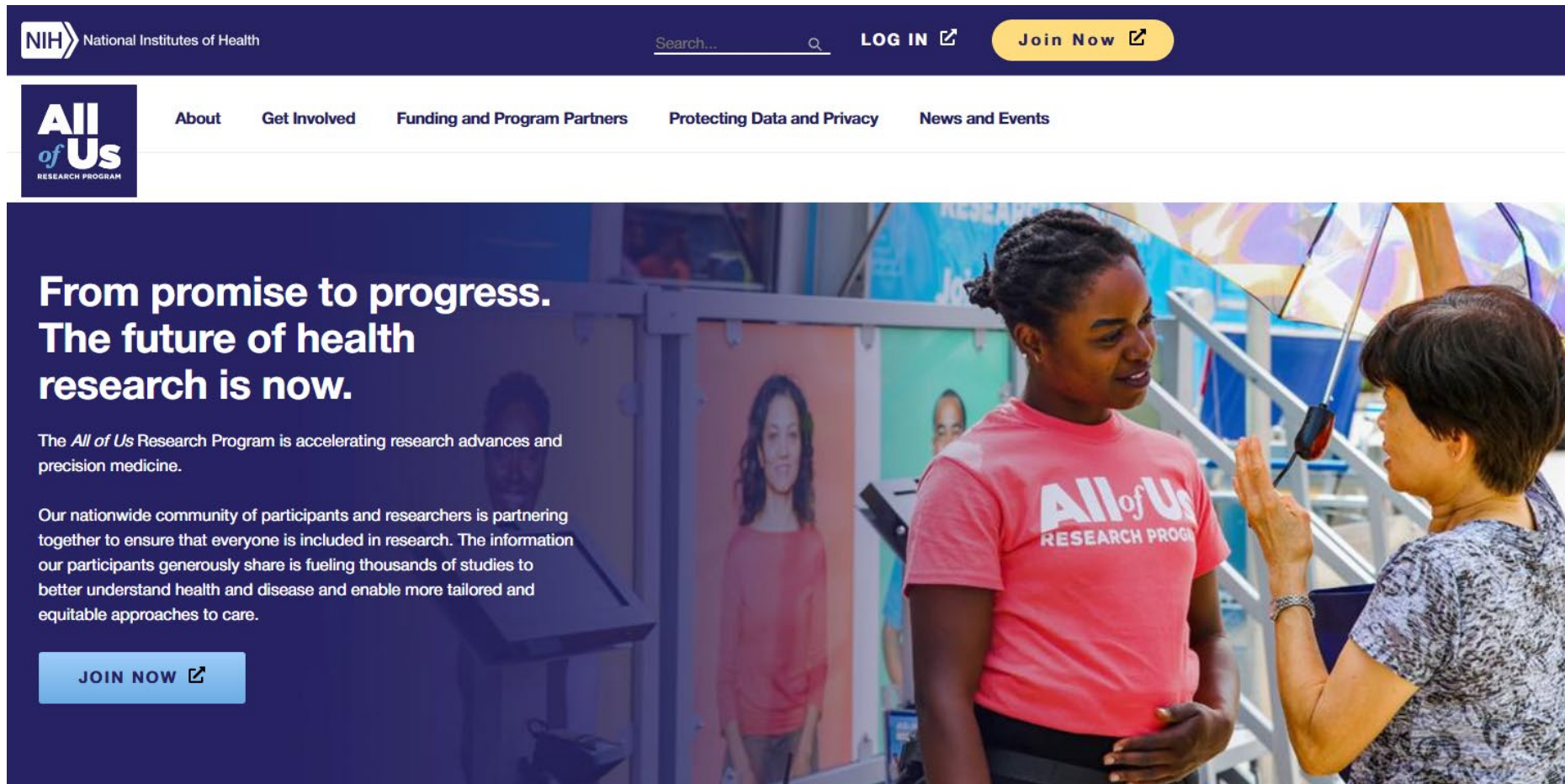
Virtual Conference

#RMED25

AOUSDOHtools:
An R Package for Social Determinants of Health Survey data in the *All of Us* Research Program

Presenter: Zhirui Deng, Biostatistician

University of Pittsburgh, School of Nursing

R/Medicine 2025 · June 13, 2025

Authors: Zhirui Deng, MS | University of Pittsburgh, School of Nursing

Theresa A. Koleck, PhD, RN | University of Pittsburgh, School of Nursing

Chen Zhang, PhD, MPH, MS | University of Rochester, School of Nursing

Peter D.R. Higgins, MD, PhD, MSc | University of Michigan, School of Medicine

Caitlin Dreisbach, PhD, RN | University of Rochester, School of Nursing

All of Us (AoU) Research Program

(All of Us Research Program Investigators, 2019)

- *All of Us*:
>1 million participant cohort

- Data:
EHR, surveys, biospecimens, wearables

- https://allofus.nih.gov/

Why Do We Need AOUSDOHtools?

- **SDOH Survey: rich data but not analysis-ready**
- **Highly disaggregated responses require transformation**
- **No built-in tools for SDOH scoring on *AoU* platform**
- **Manual scoring is time-consuming and inconsistent**
- **Need for reproducibility and accessibility**

People around here are willing to help their neighbors. (40192463)
See Answers >

People in my neighborhood generally get along with each other. (40192411)
See Answers >

People in my neighborhood can be trusted. (40192499)
See Answers >

People in my neighborhood share the same values. (40192417)
See Answers >

There is a lot of graffiti in my neighborhood. (40192420)
See Answers >

My neighborhood is noisy. (40192522)
See Answers >

People around here are willing to help their neighbors. (40192463)
See Answers ⌄

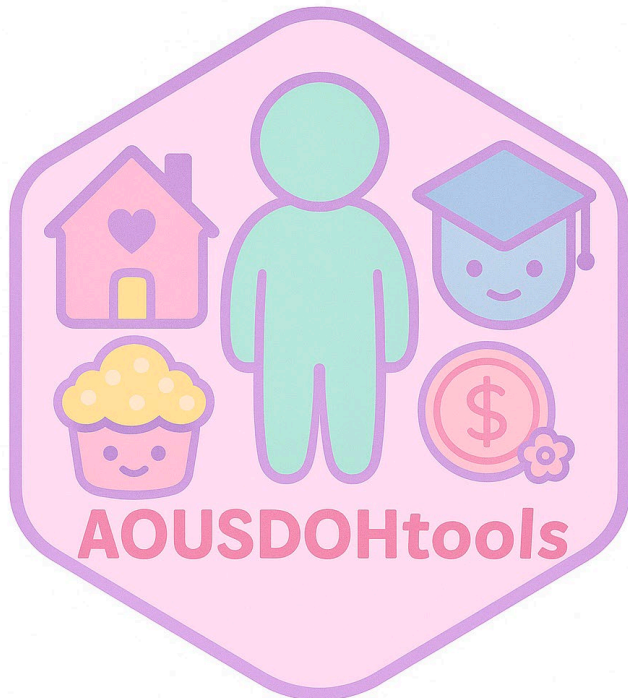| Answer	Concept Code ⓘ
Agree	40192455
Strongly agree	40192514
Neutral (neither agree nor disagree)	40192524
Disagree	40192408
Strongly disagree	40192422
Skip	903096

259,200
Participants completed this survey

80
Questions Available

What Does AOUSDOHtools Do?

- **An R package designed for *AoU* SDOH Survey users**
- **Scoring SDOH constructs based on a published user guide**
- **Koleck et al. (2024) https://pmc.ncbi.nlm.nih.gov/articles/PMC11631056/**

hexSticker package (Yu et al., 2020)

Alerts | Author Instructions | Submit | About

JAMIA

Journal of the American Medical Informatics Association

▸ J Am Med Inform Assoc. 2024 Aug 27;31(12):3032–3041. doi: 10.1093/jamia/ocae214 ↗

User guide for Social Determinants of Health Survey data in the *All of Us* Research Program

Theresa A Koleck [1,✉], Caitlin Dreisbach [2,3], Chen Zhang [4], Susan Grayson [5], Maichou Lor [6], Zhirui Deng [7], Alex Conway [8], Peter D R Higgins [9], Suzanne Bakken [10,11,12]

▸ Author information   ▸ Article notes   ▸ Copyright and License information

PMCID: PMC11631056   PMID: 39190874

Key Features and SDOH Constructs

- **30 functions across 14 SDOH constructs**
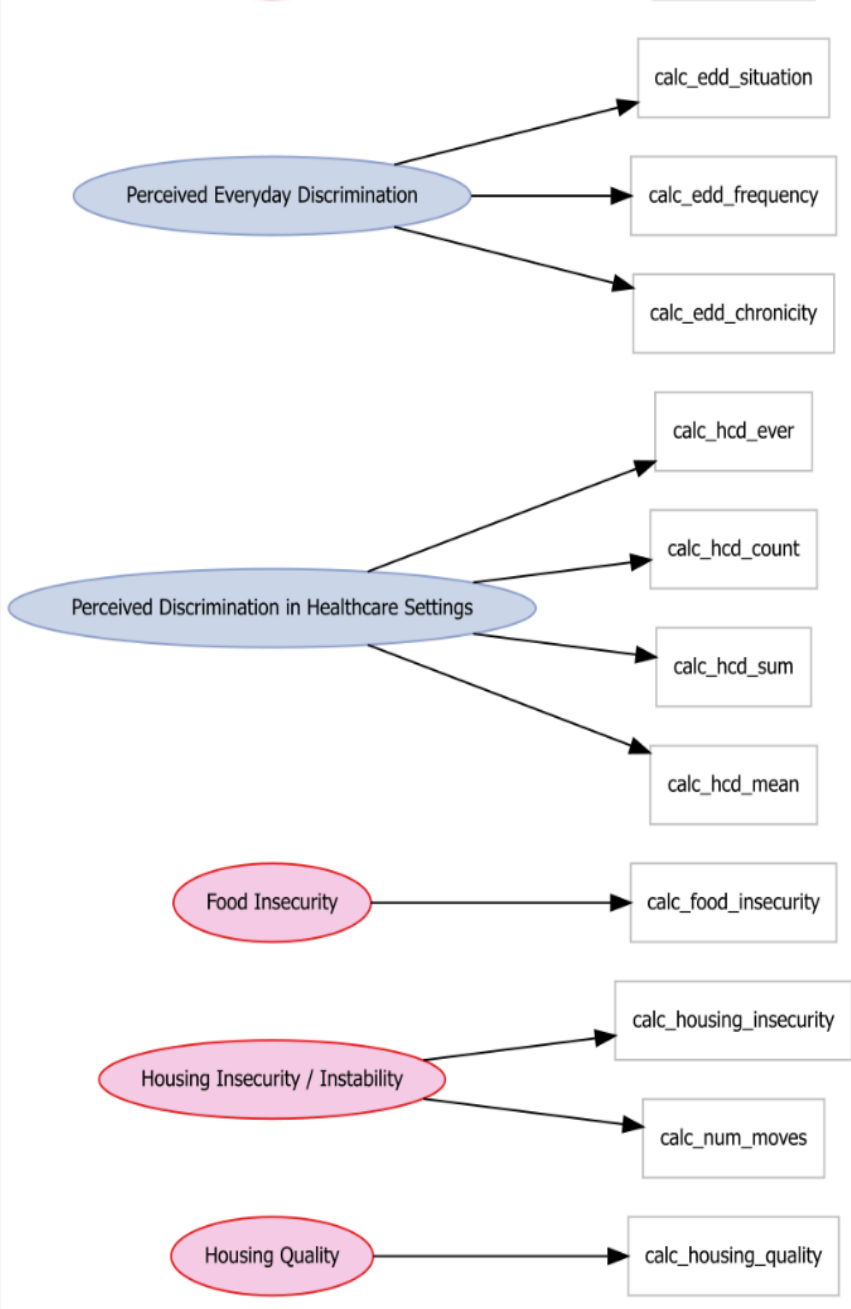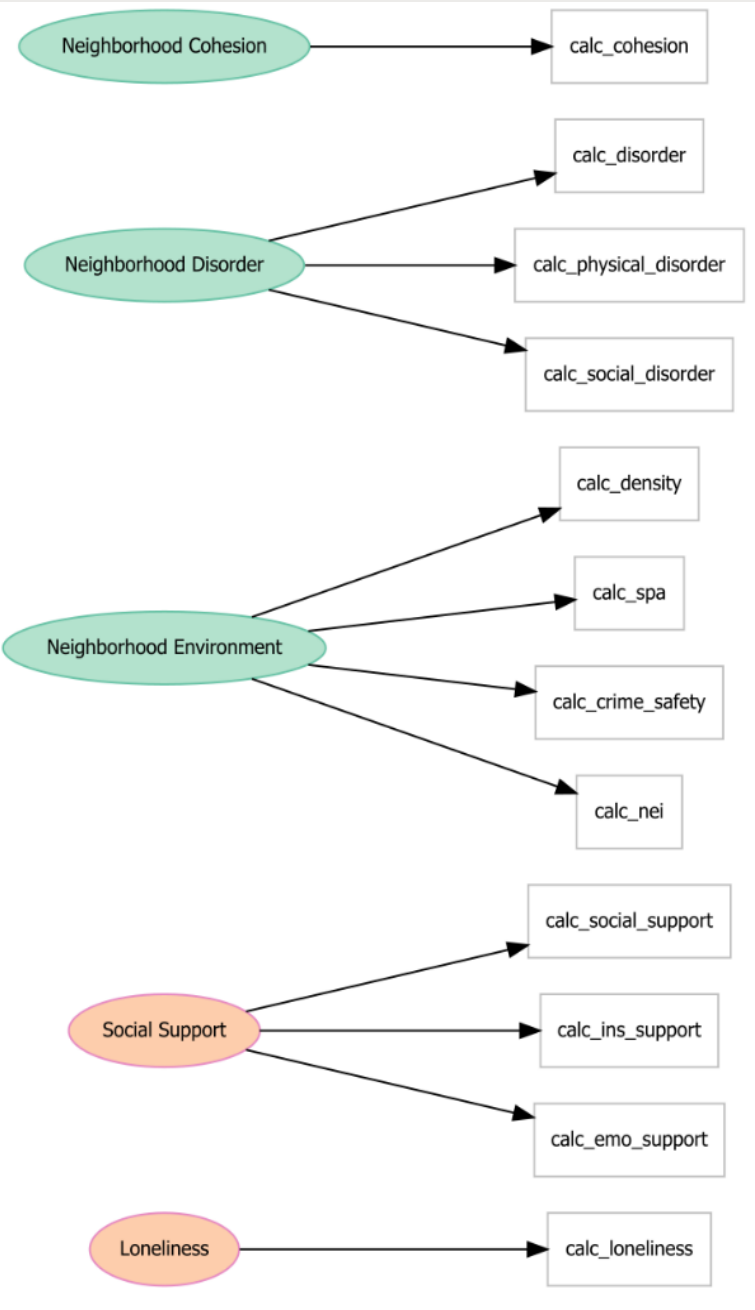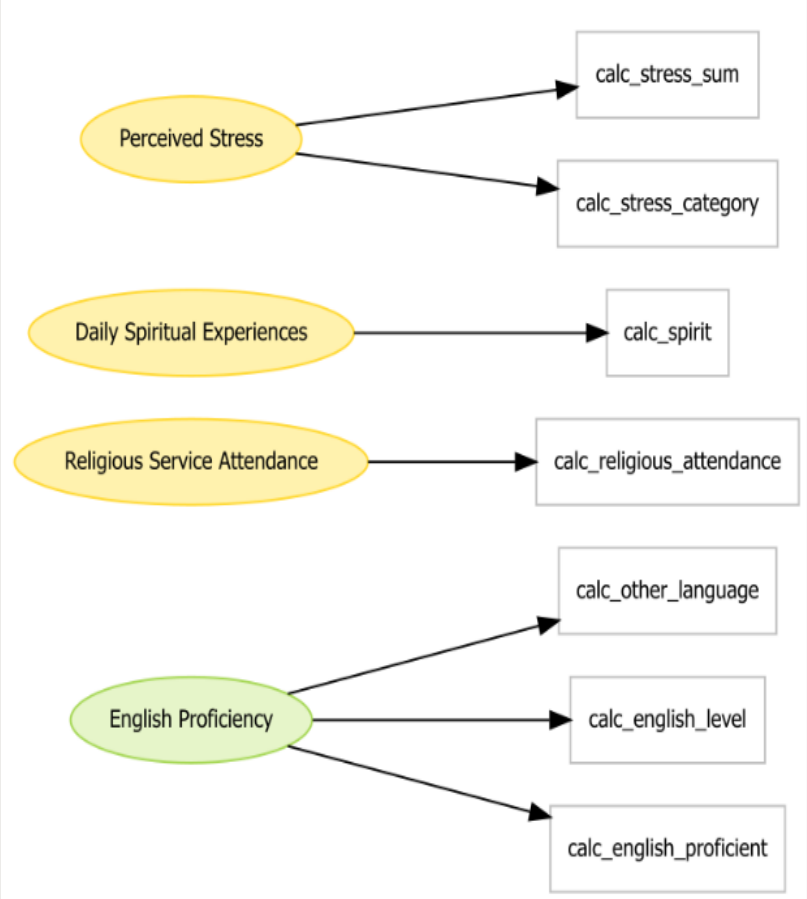- **Examples: neighborhood cohesion and neighborhood disorder**
- **Built for *All of Us* Researcher Workbench (Jupyter/RStudio environments)**

DiagrammeR package (Iannone & Roy, 2024)

14 SDOH Constructs

Constructs definitions & details can be found in Koleck et al. (2024).

All of Us Researcher Workbench

How to Use AOUSDOHtools

```
library(AOUSDOHtools)
survey_df <- your_sdoh_survey_data
cohesion_scores <- calc_cohesion(survey_df)
disorder_scores <- calc_disorder(survey_df)
……
```

```
## Create a sample survey data frame
survey_df <- data.frame(
  person_id = c(1, 1, 1, 1, 2, 2, 2, 2),
  question_concept_id = c(40192463, 40192411, 40192499, 40192417,
                          40192463, 40192411, 40192499, 40192417),
  answer_concept_id = c(40192514, 40192455, 40192524, 40192408,
                        40192514, 40192455, 40192422, 40192408)
)
```
**Example Input
(a fake dataset)**

Input:
- **SDOH Survey data includes *person_id*, *question_concept_id*, *answer_concept_id***

Output:
- **Scored datasets indexed by *person_id*, one row per participant**

Next steps:
- **Merge multiple scores (dplyr::full_join)**
- **Ready for downstream analysis**

```
head(cohesion_scores)
#> # A tibble: 2 × 2
#>   person_id cohesion
#>       <dbl>    <dbl>
#> 1         1      3.5
#> 2         2        3
```
**Example Output
(cohesion)**

calc_cohesion()

```r
calc_cohesion <- function(survey_df) {
  if (!is.null(survey_df)){
    df_cohesion  <-  survey_df |>
      dplyr::filter(question_concept_id %in% c(40192463, 40192411, 40192499, 40192417)) |> # 4 specific items
      # 40192463 = How much you agree or disagree that people around here are willing to help their neighbor?
      # 40192411 = How much you agree or disagree that people in your neighborhood generally get along with each other?
      # 40192499 = How much you agree or disagree that people in your neighborhood can be trusted?
      # 40192417 = How much you agree or disagree that people in your neighborhood share the same values?
      dplyr::select(person_id, question_concept_id, answer_concept_id) |> # map answer_concept_id to value
      dplyr::mutate(value = dplyr::case_when(
        answer_concept_id == 40192514 ~ 5, # Strongly agree
        answer_concept_id == 40192455 ~ 4, # Agree
        answer_concept_id == 40192524 ~ 3, # Neutral (neither agree nor disagree)
        answer_concept_id == 40192408 ~ 2, # Disagree
        answer_concept_id == 40192422 ~ 1, # Strongly disagree
        TRUE ~ 999)) |>
      dplyr::filter(value != 999) |> # remove skips
      dplyr::group_by(person_id) |> # group by person_id and calculate mean score
      dplyr::mutate(cohesion = round(mean(value, na.rm = TRUE), 2), # rounded to 2 decimals
                    nrows = length(value)) |> # how many questions did the participant answer?
      dplyr::filter(nrows == 4) |> # include only participants who answered all 4 questions
      dplyr::select(person_id, cohesion) |> # only 2 columns in the final result
      dplyr::distinct(person_id, .keep_all = TRUE) |> # remove duplicate rows
      dplyr::right_join(survey_df |> # include participants without scores as NA
                          dplyr::group_by(person_id) |>
                          dplyr::summarize(),
                        by = 'person_id') |>
      dplyr::ungroup() # ungroup
  }
  df_cohesion
}
```

Availability & Development

- **Package released by CRAN on 03/26/2025**
  https://r-packages.io/packages/AOUSDOHtools

Version:	1.0.4
Depends:	R (≥ 4.1.0)
Imports:	dplyr
Suggests:	testthat (≥ 3.0.0), knitr, rmarkdown, covr, devtools, usethis
Published:	2025-03-26
DOI:	10.32614/CRAN.package.AOUSDOHtools

- **Paper submitted to JOSS on 05/13/2025, currently pending**

- **Actively updating on Github**
  https://github.com/zhd52/AOUSDOHtools

Impact and Contributions

- **Standardizes SDOH score generation**

- **Simplifies data transformation**

- **Enables equity-driven research at scale**
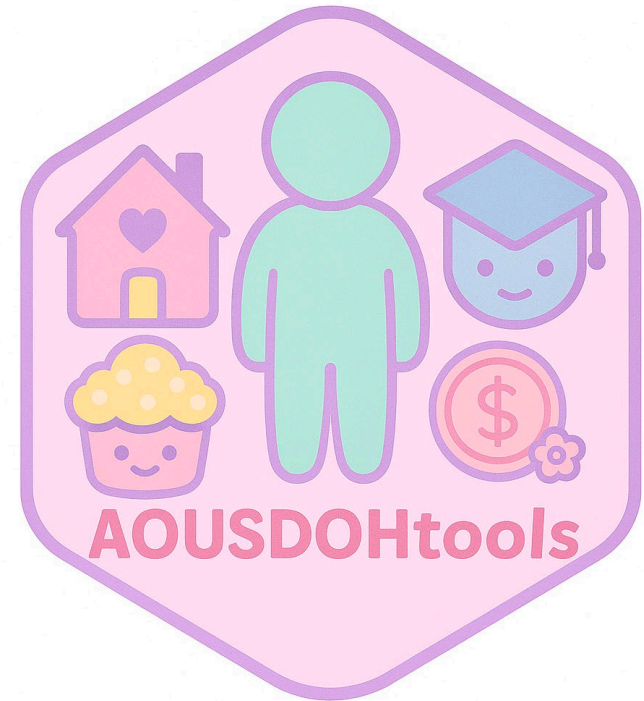
hexSticker package (Yu et al., 2020)

Get Involved

1. Register for an *All of Us* account
>   https://allofus.nih.gov/

2. Have access to *All of Us* Researcher Workbench and SDOH Survey data
>   https://www.researchallofus.org/data-tools/workbench/
>   https://www.researchallofus.org/data-tools/survey-explorer/

3. Install AOUSDOHtools
>   CRAN: install.packages("AOUSDOHtools")
>   GitHub: github.com/zhd52/AOUSDOHtools

4. Apply AOUSDOHtools to your research

Feedback and collaboration welcome!

Acknowledgements

- We thank *R/Medicine 2025* for the opportunity to present our work.

- We are grateful to the *All of Us* **Research Program and its participants** for making this work possible.

- We also acknowledge **all the authors of the published user guide (Koleck et al., 2024)**, whose work provided the conceptual and methodological foundation for scoring the Social Determinants of Health constructs implemented in AOUSDOHtools.

- We thank the **CRAN team** for supporting open-source distribution and **GitHub** for providing the platform used to develop and share the package.

References

- All of Us Research Program. *All of Us Research Program*. National Institutes of Health. Retrieved April 3, 2025, from https://allofus.nih.gov/

- All of Us Research Program Investigators. (2019). The 'All of Us' Research Program. *New England Journal of Medicine*, 381(7), 668–676. https://doi.org/10.1056/NEJMsr1809937

- All of Us Researcher Workbench. *All of Us Research Hub*. Retrieved April 3, 2025, from https://www.researchallofus.org/

- Deng, Z., Koleck, T. A., Dreisbach, C., Zhang, C., & Higgins, P. D. R. (2025). *AOUSDOHtools: Analyzing AOU SDOH Survey Data*. R package version 1.0.4. https://doi.org/10.32614/cran.package.aousdohtools

- Iannone, R., & Roy, O. (2024). *DiagrammeR: Graph/Network Visualization*. R package version 1.0.11. https://CRAN.R-project.org/package=DiagrammeR

- Koleck, T. A., Dreisbach, C., Zhang, C., Grayson, S., Lor, M., Deng, Z., Conway, A., Higgins, P. D. R., & Bakken, S. (2024). User guide for Social Determinants of Health Survey data in the All of Us Research Program. *Journal of the American Medical Informatics Association*, 31(12), 3032–3041. https://doi.org/10.1093/jamia/ocae214

- Survey Explorer – Social Determinants of Health. *All of Us Research Hub*. Retrieved April 3, 2025, from https://www.researchallofus.org/data-tools/survey-explorer/

- Wickham, H., Francois, R., Henry, L., Muller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. https://CRAN.R-project.org/package=dplyr

- Yu, G., Gatto, L., Rainer, J., Gibb, S., Corty, R., & Neitmann, T. (2020). *hexSticker: Create Hexagon Sticker in R*. R package version 0.4.9. https://CRAN.R-project.org/package=hexSticker

Useful Links

- *All of Us*: https://allofus.nih.gov/

- SDOH Survey: https://www.researchallofus.org/data-tools/survey-explorer/

- *All of Us* Researcher Workbench: https://www.researchallofus.org/data-tools/workbench/

- User Guide by Koleck et al. (2024): https://pmc.ncbi.nlm.nih.gov/articles/PMC11631056/

- CRAN: https://r-packages.io/packages/AOUSDOHtools

- Github: https://github.com/zhd52/AOUSDOHtools

Q & A

Zhirui Deng
zhd52@pitt.edu