The Journal of Open Source Software

CLPFUDatabase: A suite of R packages for energy conversion chain analysis

Matthew Kuperus Heun [1,2,3], **Zeke Marshall** [2], and **Emmanuel Aramendia** [2]

1 Engineering Department, Calvin University **2** School of Earth and Environment, University of Leeds **3** School for Public Leadership, Stellenbosch University

DOI: 10.21105/joss.06057

Software

- Review
- Repository
- Archive

Editor: Tim Tröndle

Reviewers:

- @abhishektiwari
- @nmstreethran

Submitted: 20 August 2023
Published: 09 January 2024

Summary

Energy flowing through societal energy conversion chains (ECCs) enables economic activity and facilitates human flourishing. To understand economic growth and human well-being, the field of societal energy analysis (Common, 1976) evaluates ECCs from the primary stage (resources extracted from the environment, such as coal, oil, natural gas, wind, and solar), to the final stage (energy purchased by consumers, such as refined petroleum and electricity), to the useful stage (energy desired by the end user, such as heat, motion, and light), and sometimes to energy services (such as thermal comfort, transport, and illumination). Societal *exergy* analysis (SEA) (Ertesvåg, 2001), an extension of societal *energy* analysis, quantifies ECCs in exergy terms[1].

We created a suite of open-source R packages and the metapackage CLPFUDatabase (Heun, 2023a) to assist SEA practitioners to analyze ECCs. The new packages enable analysis of any country in the world across timespans of decades or longer. In short, the new packages enable, for the first time, scalable SEA. We used the new packages to create the CL-PFU database[2], a new resource for the SEA community (Brockway et al., 2024).

This paper describes the design of the new packages and demonstrates briefly their use.

Statement of need

Historically, SEA practitioners have analyzed the ECCs of individual countries using linked spreadsheets, often starting with primary- and final-stage data from the International Energy Agency's (IEA's) Extended World Energy Balances (EWEBs) (Ayres et al., 2003; Brockway et al., 2014, 2015; Serrenho et al., 2014). Data were stored in varying and inconsistent formats. An early SEA database (De Stercke, 2014) estimated energy efficiencies of end-use machines using the economic status of countries, thereby precluding use in energy-economy studies.

The authors of the current paper and others in the field wanted to expand SEA to cover all countries, but the spreadsheet approach to SEA was deemed not scalable. A new approach to SEA was needed, one that scaled across all countries for many years without relying on economic data to estimate machine efficiencies.

[1] Exergy is the mechanical work potential of energy. See any number of internet references for an exergy primer, including ScienceDirect (2021).

[2] "CL-PFU" is an initialism for "Country-Level Primary, Final, and Useful." The CL-PFU database contains many countries as well as continental and world aggregations. Technically speaking, we create matsindf data frames stored as pins on a pinboard, not SQL or similar databases.

Design of R packages

The most important design decision for the suite of R packages involved data format. We authors are among those who developed the Physical Supply-Use Table (PSUT) framework (Aramendia et al., 2022; Guevara & Domingos, 2017; Heun et al., 2018; Rocco, 2016), an adaptation of economic Input-Output (IO) analysis (Miller & Blair, 2009) to describe energy flows through society. The Physical Supply-Use Table (PSUT) framework is a matrix-based approach that tracks energy flows ("products" in PSUT terminology) from resource extraction to processing stages ("industries" in PSUT terminology) and, ultimately, to final demand in a way that is compatible with the IEA's EWEB data (International Energy Agency, 2023). We chose the PSUT framework as the data format for the new R packages, because it succinctly describes an entire ECC for one country and one year via a set of six matrices (the "RUVY" matrices). See the following table.

Matrix	rows x columns	Name	Description
R	industry x product	Resource matrix	Contains exogeneous energy inputs to an ECC
U	product x industry	Use matrix	Describes how each energy conversion device uses energy; the sum of $\mathbf{U}_{\text{feed}}$ and $\mathbf{U}_{\text{EIOU}}$
$\mathbf{U}_{\text{feed}}$	product x industry	Feedstock use matrix	Describes feedstock inputs to energy conversion devices
$\mathbf{U}_{\text{EOIU}}$	product x industry	Energy industry own use matrix	Describes how the energy industry uses energy
V	industry x product	Make matrix	Describes how each energy conversion device makes energy
Y	product x industry	Final demand matrix	Describes how each energy carrier is consumed

Further development followed selection of the PSUT framework. First, the RUVY matrices carry the challenge that different countries and years have varying energy carriers (products) and varying energy conversion machines (industries), meaning that RUVY matrices have differing row names, differing column names, and differing sizes from one country to another and for different years within the same country. To circumvent this challenge, we created the matsbyname package (Heun, 2023b) which enables matrix mathematics that respects row and column names, inserting **0** row or column vectors when needed. Second, we wanted to be able to perform *matrix* mathematics as easily as *scalar* mathematics in R data frames using tidyverse syntax (Wickham et al., 2019). We developed the matsindf package (Heun, 2023c) to enable this functionality. Finally, manipulating row and column names proved to be a challenge, especially for RUVY matrices in matsindf data frames, so we developed the RCLabels package (Heun, 2023g) to support the PSUT framework. The table below summarizes these packages, all of which are generally useful and available on CRAN.

Package	Purpose
RCLabels	Manipulates matrix row and column names in matsindf data frames
matsbyname	Performs matrix mathematics that respects row and column names
matsindf	Stores matrices in cells of data frames, thereby enabling analyses with tidyverse syntax

Broadly speaking, four calculation steps are required to create a CL-PFU database. Each step is assisted by functions in one or more of the new R packages. First, the IEA's primary- and final-stage EWEB data must be converted to RUVY matrices for each country and year, a task accomplished by functions in the IEATools package (Heun et al., 2023). Second, human and animal muscle work must be calculated from International Labor Organization (ILO) and Food and Agriculture Organization (FAO) data, following the methodology of Steenwyk et al. (2022), using functions in the MWTools package (Marshall & Heun, 2023). The muscle work data must also be converted to RUVY matrices. Third, the IEA's primary- and final-stage ECC data must be extended to the useful stage by (a) allocating final stage energy to end-use machines and (b) multiplying allocated final energy by the final-to-useful efficiency of each machine. This task is accomplished by functions in the Recca package (Heun & Aramendia, 2023). Fourth, ECCs must be converted from energy terms to exergy terms. This step is also performed by functions in the Recca package.

The steps to create the PSUT matrices for each country and each year are accomplished by a targets (Landau, 2021) computation pipeline available in the PFUPipeline package (Heun & Marshall, 2023). A unique feature of the PFUPipeline enabled by the Recca package is an innovative exemplar system that allows analyses to proceed when allocation or efficiency data are unavailable for a country. When allocations or efficiencies are missing, a string of exemplar countries or regions are queried for the required information. For example, exemplars for Belgium are France, Europe, and ultimately, the World, in that order. The design of the calculation pipeline in the PFUPipeline package allows allocation and efficiency data for any country or year to be added at a later date to improve the database.

A second targets pipeline in the PFUAggPipeline package (Heun, 2023d) aggregates ECCs (a) by region (e.g., continents and world), (b) by energy carrier (product) category (e.g., Coal and coal products, Low-, Medium-, and High-temperature heat, etc.), (c) by energy conversion machines (industries) and/or end use sectors (e.g., Residential, Transport, etc.), and (d) to primary, final, and useful (PFU) stages. In addition, the PFUAggPipeline calculates efficiencies for each country at all available aggregations. The PFUAggPipeline uses the Recca package extensively.

Both targets pipelines (in the PFUPipeline and PFUAggPipeline packages) produce data frames readable by the pins package (Silge et al., 2023). Both pipelines benefit from the PFUSetup (Heun, 2023f) and PFUPipelineTools (Heun, 2023e) packages. PFUSetup identifies storage locations for pipeline input and output data. PFUPipelineTools provides functions and constants common to both pipelines.

The packages in the following table support creation of the CL-PFU database and are available on GitHub.

Package	Purpose
IEATools	Converts IEA EWEB data to RUVY matrices in matsindf data frames
MWTools	Converts ILO and FAO data to RUVY matrices of human and animal muscle work in matsindf data frames
Recca	Performs R energy conversion chain analysis
PFUSetup	Identifies input and output locations for the PFUPipeline and PFUAggPipeline pipelines
PFUPipelineTools	Provides basic functionality for all PFU pipelines
PFUPipeline	Provides a targets pipeline to create a data frame of RUVY matrices
PFUAggPipeline	Provides a targets pipeline to aggregate RUVY matrices

Package	Purpose
CLPFUDatabase	A metapackage (like `tidyverse`) to quickly load all package dependencies for the CL-PFU database

Example input data for the CL-PFU database (for two countries and two years) can be found in the GitHub repository for this paper. Access to the full CL-PFU database can be obtained via correspondence with author P.E.B.[3]. If desired, researchers with access to the IEA's EWEB data can create their own CL-PFU database following the example below.

Example

```r
# (0) Use the CLPFUDatabase metapackage to install all dependencies.
#     This metapackage installation step is needed only once.
devtools::install_github("EnergyEconomyDecoupling/CLPFUDatabase")

# (1) Purchase, download, and store IEA EWEB data in the correct location.
#     Sample data for two countries (Ghana and South Africa)
#     and two years (1971 and 2000) are included (with permission)
#     at the correct location in the repository for this paper.

# (2) Download FAO and ILO data for muscle work calculations
#     via the included script.
#     This download step is needed only once, and
#     downloaded data are stored in the repository
#     for this paper.
source(file.path("ExampleFolder", "DownloadScripts", "download_mw_data.R"))

# Steps (3) and (4) to construct the database are needed only once.
# Results are stored in the targets cache
# located in the "_targets" folder
# at the root level of this repository.

# (3) Run the PFUPipeline to create RUVY matrices.
targets::tar_make_future(
  script = file.path("ExampleFolder", "_targets_pfupipeline.R"),
  workers = 4
)

# (4) Run the PFUAggPipeline to aggregate
#     energy and calculate efficiencies.
targets::tar_make_future(
  script = file.path("ExampleFolder", "_targets_pfuaggpipeline.R"),
  workers = 4
)

# (5) Review results using the pins package.
#     Establish the pinboard.
pinboard <- file.path("ExampleFolder", "OutputData", "PipelineReleases") |>
  pins::board_folder(versioned = TRUE)
#     Look at the psut data frame
```

[3] Because the CL-PFU database contains primary and final energy IEA EWEB data, use of the CL-PFU database is restricted to those who have access to IEA EWEB data.

```r
#     created by the PFUPipeline.
psut <- pinboard |>
  pins::pin_read(name = "psut", version = "20231219T204403Z-2c3ce")
head(as.data.frame(psut), 2)
```

```
## Loading required package: Matrix
```

```
##   Country Method Energy.type Last.stage Year IEAMW
## 1     GHA    PCM           E      Final 1971   IEA
## 2     GHA    PCM           E      Final 2000   IEA
##                                                  R
## 1 <S4 class 'dgCMatrix' [package "Matrix"] with 6 slots>
## 2 <S4 class 'dgCMatrix' [package "Matrix"] with 6 slots>
##                                                  U
## 1 <S4 class 'dgCMatrix' [package "Matrix"] with 6 slots>
## 2 <S4 class 'dgCMatrix' [package "Matrix"] with 6 slots>
##                                             U_feed
## 1 <S4 class 'dgCMatrix' [package "Matrix"] with 6 slots>
## 2 <S4 class 'dgCMatrix' [package "Matrix"] with 6 slots>
##                                             U_EIOU
## 1 <S4 class 'dgeMatrix' [package "Matrix"] with 4 slots>
## 2 <S4 class 'dgeMatrix' [package "Matrix"] with 4 slots>
##                                             r_EIOU
## 1 <S4 class 'dgeMatrix' [package "Matrix"] with 4 slots>
## 2 <S4 class 'dgeMatrix' [package "Matrix"] with 4 slots>
##                                                  V
## 1 <S4 class 'dgCMatrix' [package "Matrix"] with 6 slots>
## 2 <S4 class 'dgCMatrix' [package "Matrix"] with 6 slots>
##                                                  Y
## 1 <S4 class 'dgCMatrix' [package "Matrix"] with 6 slots>
## 2 <S4 class 'dgCMatrix' [package "Matrix"] with 6 slots>
##                                            S_units
## 1 <S4 class 'dgeMatrix' [package "Matrix"] with 4 slots>
## 2 <S4 class 'dgeMatrix' [package "Matrix"] with 4 slots>
```

```r
#     Each column of the psut data frame contains sparse matrices,
#     the columns of which are suppressed on display.
#     Show the column names here.
colnames(psut$R[[1]])
```

```
##  [1] "Aviation gasoline"
##  [2] "Crude oil"
##  [3] "Electricity"
##  [4] "Fuel oil"
##  [5] "Gas/diesel oil excl. biofuels"
##  [6] "Hydro [from Resources]"
##  [7] "Kerosene type jet fuel excl. biofuels"
##  [8] "Lubricants"
##  [9] "Other kerosene"
## [10] "Primary solid biofuels [from Resources]"
```

```r
#     Show an example resource (R) matrix.
psut$R[[1]]
```

```
## 9 x 10 sparse Matrix of class "dgCMatrix"
```

```
##   [[ suppressing 10 column names 'Aviation gasoline', 'Crude oil', 'Electricity' ...
```

```
##
```

```
## Imports [of Aviation gasoline]                    44.7988        .      .
## Imports [of Crude oil]                              .      38359.8007 .
## Imports [of Kerosene type jet fuel excl. biofuels]  .          .      .
## Imports [of Lubricants]                             .          .      .
## Imports [of Other kerosene]                         .          .      .
## Resources [of Hydro]                                .          .      .
## Resources [of Primary solid biofuels]               .          .      .
## Statistical differences                             .          0.0001 0.0084
## Stock changes [of Gas/diesel oil excl. biofuels]    .          .      .
##
## Imports [of Aviation gasoline]                    .        .          .
## Imports [of Crude oil]                            .        .          .
## Imports [of Kerosene type jet fuel excl. biofuels] .       .          .
## Imports [of Lubricants]                           .        .          .
## Imports [of Other kerosene]                       .        .          .
## Resources [of Hydro]                              .        .      10472.4
## Resources [of Primary solid biofuels]             .        .          .
## Statistical differences                            1e-04 822.6979     .
## Stock changes [of Gas/diesel oil excl. biofuels]  .      476.2987     .
##
## Imports [of Aviation gasoline]                      .        .        .
## Imports [of Crude oil]                              .        .        .
## Imports [of Kerosene type jet fuel excl. biofuels] 892.0019  .        .
## Imports [of Lubricants]                             .      755.9979  .
## Imports [of Other kerosene]                         .        .      43.7981
## Resources [of Hydro]                                .        .        .
## Resources [of Primary solid biofuels]               .        .        .
## Statistical differences                             .        .      0.0042
## Stock changes [of Gas/diesel oil excl. biofuels]    .        .        .
##
## Imports [of Aviation gasoline]                        .
## Imports [of Crude oil]                                .
## Imports [of Kerosene type jet fuel excl. biofuels]    .
## Imports [of Lubricants]                               .
## Imports [of Other kerosene]                           .
## Resources [of Hydro]                                  .
## Resources [of Primary solid biofuels]             87400
## Statistical differences                               .
## Stock changes [of Gas/diesel oil excl. biofuels]      .
```

```r
#     Review the aggregate PFU efficiency data frame
#     created by the PFUAggPipeline.
agg_eta_pfu <- pinboard |>
  pins::pin_read(name = "agg_eta_pfu", version = "20231211T161827Z-8f5e7")
agg_eta_pfu |>
#     Filter and delete columns
#     to present a subset of exergy results in a data frame.
#     The EX.p, EX.f, and EX.u columns contain aggregate exergy at
#     primary, final, and useful stages, respectively, in TJ.
#     The eta_pf, eta_fu, and eta_pu columns contain aggregate
#     primary-to-final, final-to-useful, and
#     primary-to-useful efficiencies.
  dplyr::filter(Energy.type == "X", IEAMW == "Both",
              Product.aggregation == "Specified",
              Industry.aggregation == "Specified",
```

```
                 GrossNet == "Net",
                 Country %in% c("GHA", "ZAF")) |>
  dplyr::mutate(
    Method = NULL,
    Energy.type = NULL,
    Product.aggregation = NULL,
    Industry.aggregation = NULL,
    IEAMW = NULL,
    Chopped.mat = NULL,
    Chopped.var = NULL,
    GrossNet = NULL
  )
```

```
## # A tibble: 8 × 9
##   Country Last.stage  Year     EX.p     EX.f    EX.u eta_pf eta_fu eta_pu
##   <chr>   <chr>      <dbl>    <dbl>    <dbl>   <dbl>  <dbl>  <dbl>  <dbl>
## 1 GHA     Final       1971  179666.  139694.  16853.  0.778  0.121 0.0938
## 2 GHA     Useful      1971  179666.  139694.  16853.  0.778  0.121 0.0938
## 3 GHA     Final       2000  344825.  274173.  56749.  0.795  0.207 0.165
## 4 GHA     Useful      2000  344825.  274173.  56749.  0.795  0.207 0.165
## 5 ZAF     Final       1971 2138435. 1525335. 211670.  0.713  0.139 0.0990
## 6 ZAF     Useful      1971 2138435. 1525335. 211670.  0.713  0.139 0.0990
## 7 ZAF     Final       2000 4957435. 2451131. 482533.  0.494  0.197 0.0973
## 8 ZAF     Useful      2000 4957435. 2451131. 482533.  0.494  0.197 0.0973
```

Conclusion

The new open source R packages described in this paper enable, for the first time, creation of databases of primary, final, and useful energy and exergy conversion chains in a scalable manner. Such databases will enable advances in societal exergy analysis and lead to new insights about the role of energy in economic growth and human well-being. An example of early work using the CL-PFU database is the calculation of useful stage Energy Return On Investment (EROI) for fossil fuels (Aramendia et al., 2023), for which the ECCTools (Aramendia & Heun, 2022a) and EROITools (Aramendia & Heun, 2022b) packages were developed, expanding the software capabilities introduced here[4].

Acknowledgements

M.K.H. acknowledges the support of Calvin University for a sabbatical during which he developed the PSUT framework and course releases during which he wrote many of the R packages described in this paper. Z.M.'s time was funded by the UK Research Council under EPSRC Fellowship award EP/R024254/1. E.A.'s time was funded by the University of Leeds (School of Earth and Environment). M.K.H., Z.M., and E.A. are grateful for support and encouragement from Dr. Paul E. Brockway (University of Leeds). It was Dr. Brockway's EPSRC fellowship that provided the need for the software described in this paper.

References

Aramendia, E., Brockway, P., Taylor, P., Norman, J., Heun, M. K., & Marshall, Z. (2023). *Chapter 4: Estimation of fossil fuels useful stage energy return on investment and implica-*

[4]Note that the ECCTools and EROITools packages are not the subject of this paper and were not peer reviewed.

tions for renewable energy systems in Exploring the useful energy implications of the global energy transition: A net energy perspective [PhD thesis]. University of Leeds.

Aramendia, E., & Heun, M. K. (2022a). *ECCTools: Tools for modifying the energy conversion chain*. https://doi.org/10.5281/zenodo.5841962

Aramendia, E., & Heun, M. K. (2022b). *EROITools: Tools for calculating and aggregating energy return on investment values*. https://doi.org/10.5281/zenodo.7696869

Aramendia, E., Heun, M. K., Brockway, P. E., & Taylor, P. G. (2022). Developing a multi-regional physical supply use table framework to improve the accuracy and reliability of energy analysis. *Applied Energy*, *310*(118413), 1–20. https://doi.org/10.1016/j.apenergy.2021.118413

Ayres, R. U., Ayres, L. W., & Warr, B. (2003). Exergy, power, and work in the US economy, 1900–1998. *Energy*, *28*(3), 219–273. https://doi.org/10.1016/s0360-5442(02)00089-0

Brockway, P. E., Barrett, J. R., Foxon, T. J., & Steinberger, J. K. (2014). Divergence of trends in US and UK aggregate exergy efficiencies 1960–2010. *Environmental Science & Technology*, *48*, 9874–9881. https://doi.org/10.1021/es501217t

Brockway, P. E., Heun, M. K., Marshall, Z., Aramendia, E., Steenwyk, P., Relph, T., Widjanarko, M., Kim, J. (James), Sainju, A., & Irtube, J. (2024). A country-level primary-final-useful (CL-PFU) energy and exergy database: Overview of its construction and 1971–2020 world-level efficiency results. *Environmental Research: Energy*, (In review).

Brockway, P. E., Steinberger, J. K., Barrett, J. R., & Foxon, T. J. (2015). Understanding China's past and future energy demand: An exergy efficiency and decomposition analysis. *Applied Energy*, *155*, 892–903. https://doi.org/10.1016/j.apenergy.2015.05.082

Common, M. (1976). The economics of energy analysis reconsidered. *Energy Policy*, *4*(2), 158–165. https://doi.org/10.1016/0301-4215(76)90008-2

De Stercke, S. (2014). *Dynamics of energy systems: A useful perspective* (Interim Report IR-14-013). IIASA.

Ertesvåg, I. S. (2001). Society exergy analysis: A comparison of different societies. *Energy*, *26*, 253–270. https://doi.org/10.1016/s0360-5442(00)00070-0

Guevara, Z., & Domingos, T. (2017). The multi-factor energy input–output model. *Energy Economics*, *61*, 261–269. https://doi.org/10.1016/j.eneco.2016.11.020

Heun, M. K. (2023a). *CLPFUDatabase: Easily install and load packages to construct the CL-PFU database*. https://doi.org/10.5281/zenodo.10359832

Heun, M. K. (2023b). *matsbyname: An implementation of matrix mathematics*. https://doi.org/10.5281/zenodo.5118872

Heun, M. K. (2023c). *matsindf: Matrices in data frames*. https://doi.org/10.5281/zenodo.5475443

Heun, M. K. (2023d). *PFUAggPipeline: Calculate aggregations and efficiencies for energy conversion chains in PSUT format*. https://doi.org/10.5281/zenodo.6409759

Heun, M. K. (2023e). *PFUPipelineTools: Tools for PFU pipelines*. https://doi.org/10.5281/zenodo.8226419

Heun, M. K. (2023f). *PFUSetup: Automated setup for the PFU database project*. https://doi.org/10.5281/zenodo.5228359

Heun, M. K. (2023g). *RCLabels: Tools for munging and manipulating row and column labels on matrices*. https://doi.org/10.5281/zenodo.5819143

Heun, M. K., & Aramendia, E. (2023). *Recca: R energy conversion chain analysis*. https://doi.org/10.5281/zenodo.5226085

Heun, M. K., Aramendia, E., & Marshall, Z. (2023). *IEATools: Tools for munging and manipulating IEA extended energy balance data*. https://doi.org/10.5281/zenodo.5086371

Heun, M. K., & Marshall, Z. (2023). *PFUPipeline: Functions and workflows for primary-final-useful societal exergy analysis within the PSUT framework*. https://doi.org/10.5281/zenodo.5228375

Heun, M. K., Owen, A., & Brockway, P. E. (2018). A physical supply-use table framework for energy analysis on the energy conversion chain. *Applied Energy*, *226*, 1134–1162. https://doi.org/10.1016/j.apenergy.2018.05.109

International Energy Agency. (2023). *IEA World Extended Energy Balances*. https://www.iea.org/data-and-statistics/data-product/world-energy-balances.

Landau, W. M. (2021). The targets R package: A dynamic make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, *6*(57), 2959. https://doi.org/10.21105/joss.02959

Marshall, Z., & Heun, M. K. (2023). *MWTools: Functions for calculation of muscle work within the PSUT framework*. https://doi.org/10.5281/zenodo.7584858

Miller, R. E., & Blair, P. D. (2009). *Input-output analysis: Foundations and extensions* (2nd ed.). Cambridge University Press.

Rocco, M. V. (2016). *Primary exergy cost of goods and services: An input–output approach*. Springer.

ScienceDirect. (2021). *Exergy*. https://www.sciencedirect.com/topics/earth-and-planetary-sciences/exergy.

Serrenho, A. C., Sousa, T., Warr, B., Ayres, R. U., & Domingos, T. (2014). Decomposition of useful work intensity: The EU (European Union)-15 countries from 1960 to 2009. *Energy*, *76*, 704–715. https://doi.org/10.1016/j.energy.2014.08.068

Silge, J., Wickham, H., & Luraschi, J. (2023). *Pins: Pin, discover, and share resources*. https://CRAN.R-project.org/package=pins

Steenwyk, P., Heun, M. K., Brockway, P., Sousa, T., & Henriques, S. (2022). The contributions of muscle and machine work to land and labor productivity in world agriculture since 1800. *Biophysical Economics and Sustainability*, *7*(2), 1–17. https://doi.org/10.1007/s41247-022-00096-z

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686