

IMPROVING OPEN DATA ACCESSIBILITY THROUGH PACKAGE DEVELOPMENT AND COMMUNITY WORK

Diego Kozlowski¹ Pablo Tiscornia² Guido Weksler^{3,4} German Rosati^{4,5} Natsumi Shokida⁶ Antonio Vazquez Brust^{7,8} Demian Zayat⁹

Elio Campitelli^{10,4}

¹ FSTM-UL, ² INDEC, ³ FCE-UBA, ⁴ CONICET, ⁵ IHSS-UNSAM, ⁶ Economía Femeni(s)ta, ⁷ FLACSO, ⁸ UTDT, ⁹ FD-UBA, ¹⁰ CIMA

2020



The **eph** and **presentes** packages were developed by members of the R user group (RUG) in Buenos Aires, *RenBaires*, involving developers with diverse backgrounds. Their purpose is to improve access to public information.

It is an example of the impact of having a strong regional community on package development for improving data access.

EPH - holatam.github.io/eph

OVERVIEW

The **eph** package [1] has as a goal to facilitate the work of those R-users that work with the Argentinian Permanent Household survey, which doesn't count with an official API. Some of its functionalities are:

- ▶ Data gathering,
- ▶ building data pools for cross-time analysis,
- ▶ Organize the information from nomenclatures of occupation and economic activities,
- ▶ organize labels of the database,
- ▶ map the information by agglomerates,
- ▶ replicate the official methodology for poverty and indigence.

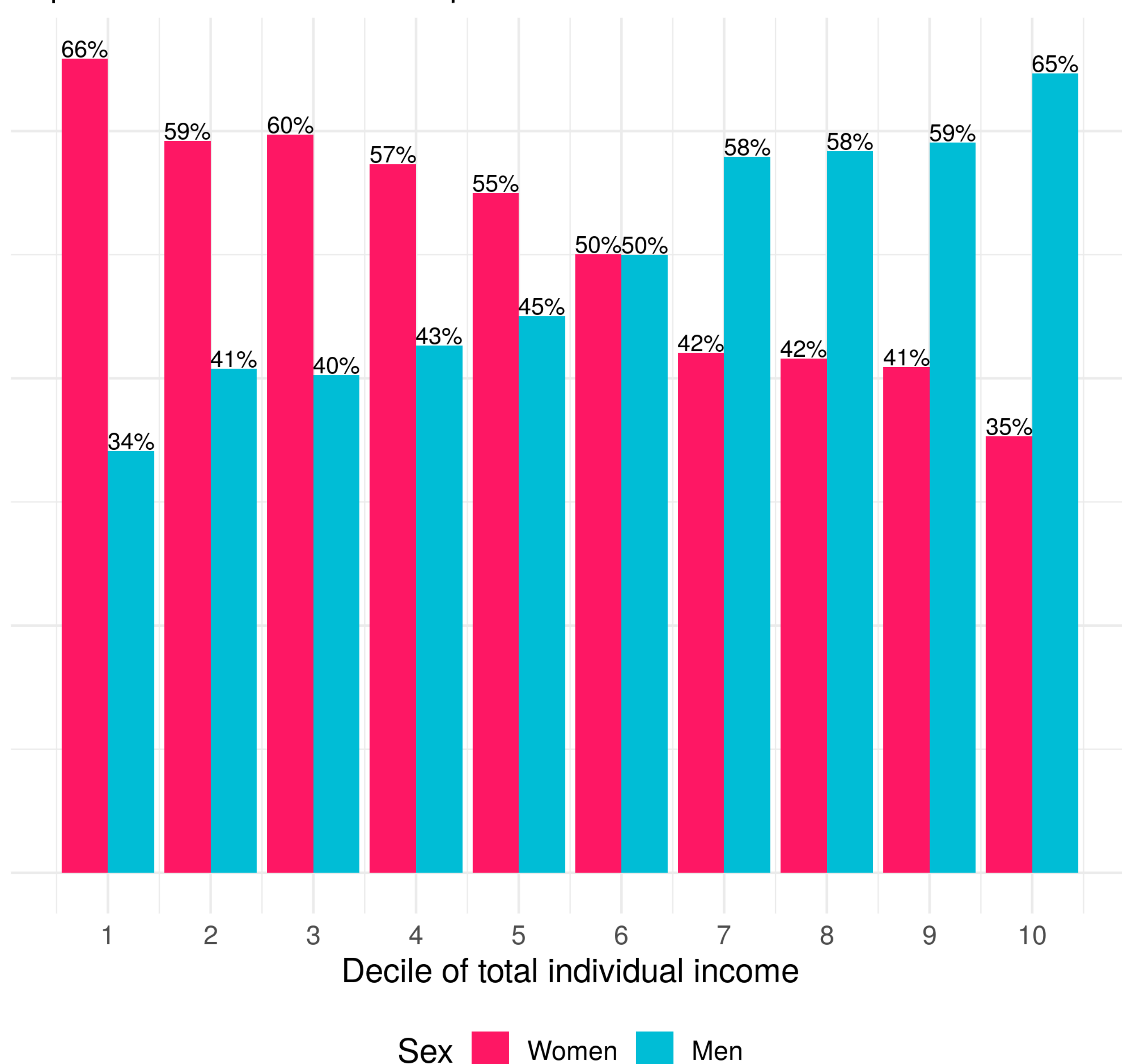
GOALS

- ▶ With this, we aim to ease the work of non-expert users, so they can focus on the data analysis, instead of the technical details. We also include warnings and detailed documentation for raising awareness on those things that might have an impact on the results (like data validity).
- ▶ As the majority of the users of the survey come from Argentina or elsewhere in Latin America, and as a way to bring the R code towards our community, the documentation of the package was written in Spanish.

One use-case of this package is in [3], a periodical report on gender inequalities in Argentina. Figure 1 was taken from this report.

FIGURE 1

Sex composition of the deciles of total individual income
Population with incomes. 3rd quarter 2019



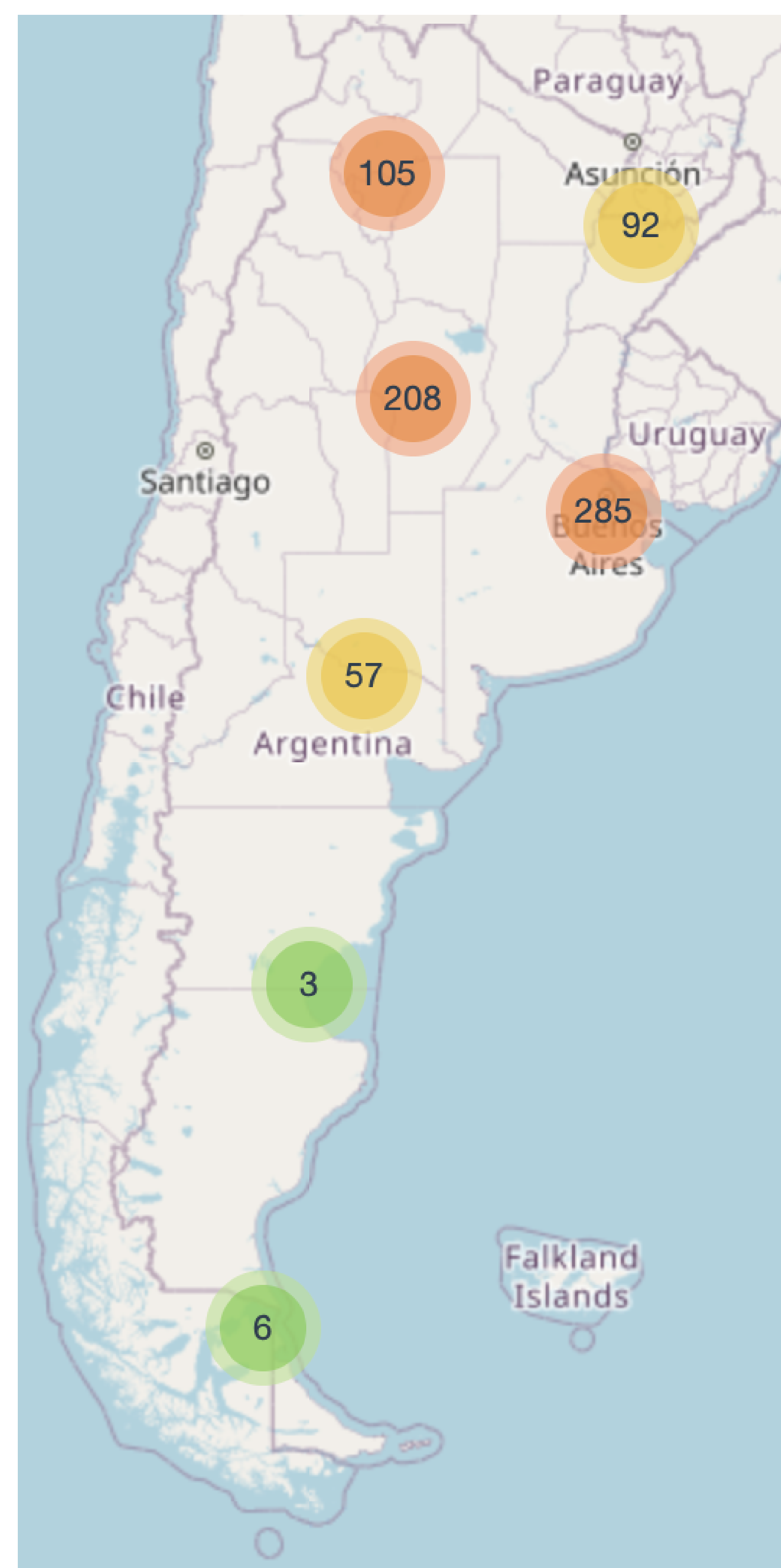
presentes. diegokoz.github.io/presentes

OVERVIEW

The **presentes** package includes the publicly available data about victims of state terrorism during the last military dictatorship in Argentina. The extensive research made by the unique registry of victims of state terrorism (RUVTE) and the memory park is available mostly in PDF files [4] or a webpage not available for bulk download [5]. And include information about:

- ▶ Official victims of illegal repression,
- ▶ victims of illegal repression without a legal claim,
- ▶ victims data from the *Parque de la memoria* (memory park) database,
- ▶ clandestine detention centers.

FIGURE 2



GOALS

These dataset include many relevant personal information about the victims origin as well as place and date of detention/kidnapping and discovery of the mortal remains. Also, Clandestine Detention Center (CDC) records where extended with geolocalization obtained from their addresses. Figure 2 shows the location of the CDC using Leaflet [6].

ACKNOWLEDGEMENT

The Doctoral Training Unit **Data-driven computational modelling and applications** (DRIVEN) is funded by the Luxembourg National Research Fund under the PRIDE programme (PRIDE17/12252781). <https://driven.uni.lu>

REFERENCES

- [1] Diego Kozlowski et al. *holatam/eph: dplyr compatibilities*. Version 0.3.1. May 2020. doi: 10.5281/zenodo.3842011. URL: <https://doi.org/10.5281/zenodo.3842011>.
- [2] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: <https://ggplot2.tidyverse.org>.
- [3] Natsumi Shokida, Daiana Serpa, and Julieta Moure. *La desigualdad de género se puede medir*. URL: https://ecofeminita.github.io/EcoFemiData/informe_desigualdad_genero/trim_2019_03/informe.nb.html (visited on 06/08/2020).
- [4] RUVTE. *Informe de Investigación*. es. Oct. 2017. URL: <https://www.argentina.gob.ar/sitiosdememoria/ruvte/informe> (visited on 06/08/2020).
- [5] Parque de la Memoria. *Base de datos de consulta pública*. es-ES. URL: <http://basededatos.parquedelamemoria.org.ar/registros/> (visited on 06/08/2020).
- [6] Joe Cheng, Bhaskar Karambelkar, and Yihui Xie. *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. R package version 2.0.3. 2019. URL: <https://CRAN.R-project.org/package=leaflet>.