Improving open data accessibility through package development and community work

Diego Kozlowski[1]   Pablo Tiscornia[2]   Guido Weksler[3,4]
German Rosati[4,5]   Natsumi Shokida[6]   Antonio Vazquez Brust[7,8]
Demian Zayat[9]   Elio Campitelli[10,4]

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

2020

Motivation

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

It is an example on how a strong regional community can help development of packages and improve data access.

EPH - overview

The **eph** [1] package (Kozlowski et al. 2020) has as a goal to facilitate the work of those R-users that use the Argentian Permanent Household survey, which doesn't count with an official API. Some of its functionalities are:

- ▶ Data gathering,
- ▶ build 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, and
- ▶ replicate the official methodology for poverty and indigence.
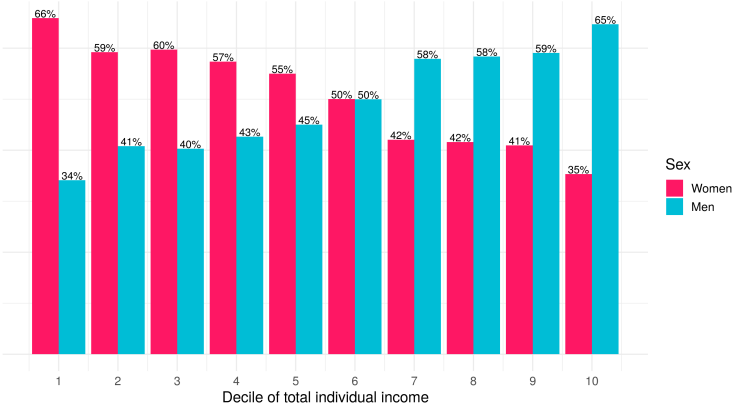
---

[1] holatam.github.io/eph

EPH - goals

- ▶ 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 is in Spanish.

EPH - example of use

As an example of use, Shokida, Serpa, and Moure 2020 produce a periodical report on gender inequalities in Argentina. The following figure was taken from this report.

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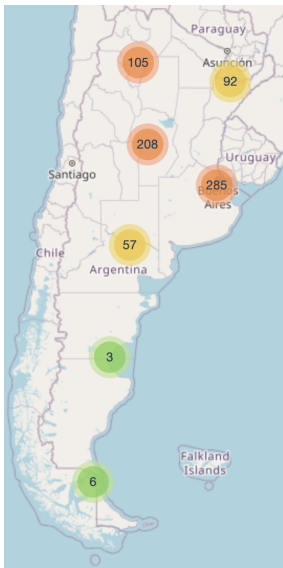Population with incomes. 3rd quarter 2019

presentes - overview

The **presentes** [2] package includes the public data about victims of state terrorism during the last military dictatorship in Argentina.
The extensive research made by the Unique Registry of State-Terrorism' Victims (RUVTE 2017) and the Memory Park (Memoria 2020) is not available in a database-format. They share information about:

► Victims of illegal repression with and without legal claim, and
► Clandestine Detention Centers (CDC).

---

[2]diegokoz.github.io/presentes

These datasets include many relevant personal information about the victims origin as well as their place & date of detention and the discovery of their mortal remains. We also extended the CDC records with geolocatization obtained from their addresses. The figure shows the distribution of the CDC using Leaflet (Cheng, Karambelkar, and Xie 2019).

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).

Further reading

[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] 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).

[3] RUVTE. *Informe de Investigación*. es. Oct. 2017. URL: https://www.argentina.gob.ar/sitiosdememoria/ruvte/informe (visited on 06/08/2020).

[4] Parque de la Memoria. *Base de datos de consulta pública*. es-ES. URL: http://basededatos.parquedelamemoria.org.ar/registros/ (visited on 06/08/2020).

[5] 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.