Divide and Conquer: From polar To the Polarverse

Juan Pablo Ruiz Nicolini

Published December 5, 2023

Citation

Keywords
Community, Packages, Development

Copyright
Copyright © Juan Pablo Ruiz Nicolini 2023. Distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
"I'm just a political scientist, standing in front of the R console asking it to help me finish my thesis."

It was 2015, a Political Science master’s student had to process data to hand in his thesis and decided to take the opportunity to learn how to use R. To the long and winding road of the academic requirement was added an extra degree of difficulty: incorporating programming software from scratch, with a somewhat steep learning curve.

Years later, as I review scripts from the thesis code, I smile to see long lines of code to recode variables with a structure like this:

```r
#PARTY LABELS A GOBERNADOR# 0 VOTOS EN BLANCO# 1 WAYAR GOBERNADOR (WALTER WAYAR)# 2 COALICION CIVICA (ARI) (ALBERTO ROLANDO TONDA)# 3 FRENTE JUSTICIALISTA RENOVADOR DE LA VICOTRIA (JUAN MANUEL URTUBEY)# 4 FRENTE OLMEDO GOBERNADOR (ALFREDO OLMEDO)# 5 MORELLO GOBERNADOR / PROYECTO SUR (CARLOS FERNANDO MORELLO)# 6 PARTIDO OBRERO (CLAUDIO DEL PLA)# 7 UNION CIVICA RADICAL (DIEGO MARINO)# 8 MIJD (CARLOS ARIEL FERNANDEZ)#RECODE TODOS LOS PARTY LABELS SEGUN ADHESIONES A LISTAS CON CANDIDATOS A GOBERNADORSALTA$ListaF[ SALTA$Lista == "VOTOS EN BLANCO" & SALTA$Year == 2011] <-9999SALTA$ListaF[ SALTA$Lista == "AGRUPACION CRECIMIENTO SOLIDARIO" & SALTA$Year == 2011] <-3SALTA$ListaF[ SALTA$Lista == "AGRUPACION DEMOCRATICA COMUNAL" & SALTA$Year == 2011] <-3SALTA$ListaF[ SALTA$Lista == "AGRUPACION DEMOCRATICA TARTAGAL (ADETA)" & SALTA$Year == 2011] <-3SALTA$ListaF[ SALTA$Lista == "AGRUPACION FEDERALISTA POPULAR MOSCONI" & SALTA$Year == 2011] <-3
```

Looking at it in perspective (like my scripts) I can say that learning R was the best decision I made in my professional career.

R Since then it has become a fundamental tool in my daily life: As academic in my work as a consultant and analyst; in my role as a public servant; in my spare time, as a hobby; and in my current role, coordinating the Data team at Fundar, a think tank that studies and formulates public policies for the sustainable and inclusive development of Argentina.

Of all the qualities that R has, the community was the determining factor in my journey with this programming language.

While I was finishing the first step of the thesis, I found out about the organization of the first Latin American Conference on R for R&D (LatinR) to which I owe my second (and third... and umpteenth) step in my relationship with R. In the first conference I presented use cases for political analyses from R, implementing grids from Argentina to use with ggplot as if they were maps with geofacet. In the second conference, we
presented the Shiny App **Electoral Intelligence** for the analysis of electoral results in Argentina and I was able to participate in the Package Development workshop with Hadley Wickham. This was the ideal starting point to get started with the idea of packaging code myself.

geofacetAR R package hex logo on the left and original polAr R package hex logo on the right

### The birth of polAr

With the aim of presenting at the third LatinR conference, I developed a polished version of geofacetAR with functions to work with the design of grids as if they were maps of Argentina.

The Covid19 pandemic hit. Learning how to develop packages, in times of confinement and many hours in front of the computer, turned the electoral data processing work of my life as a political scientist and consultant into a new package: polAr; that on May 16, 2020 traveled to CRAN.

### The polAr Universe

Although originally the package sought to facilitate access and tools for the analysis of electoral data in Argentina, over time, other workflows and data sources (geographic, presidential speeches, and legislative votes) were integrated, with an increase in dependencies and difficulty in managing more and more functions. This was a problematic issue that was discussed with colleagues in the Package Development panel during the third LatinR conference (2020).

polAr had reached 4000 downloads when I received an email from the CRAN team informing me that I had to fix a policy violation to prevent the package from being archived. Unfortunately my submitted fix did not appear to correct the problem and the package was archived.

### The polarverse

If the conversation at LatinR made me think about splitting polAr into multiple packages as a possibility, the CRAN archiving gave me the opportunity to do so. The decision was to generate independent workflows; a strategy that made it possible to move forward at different paces, making specific decisions for each topic, trying to address it at the same time as part of a whole, emulating the Tidyverse paradigm: a similar philosophy and grammar of operation for the collection of packages. The goal was to improve the development of that which led to the archiving of polAr.
I presented “Divide and Conquer: from \{polAr\} to the polarverse” at the fourth edition of LatinR, showcasing this new universe with five packages:

1. **opinAr**: provides access to data on public opinion and government in Argentina based on the Confidence in Government Index published monthly by the Universidad Torcuato Di Tella.
2. **electorAr**: facilitates access to electoral data for Argentina (for different levels and with different spatial aggregation) and functions to compute relevant indicators.
3. **legislAr**: facilitates downloading disaggregated information of votes in the legislative chambers of Argentina, originating from the project Decade Voted project by Andy Tow.
4. **discursAr**: is designed to obtain data on political speeches. In principle, it provides access to presidential speeches in legislative assemblies at the inauguration of ordinary sessions.
5. **geoAr**: allows access to data and tools for spatial workflows from within R. One paper on the package was accepted for the sixth edition of LatinR conference (2023).

In addition, we worked on an auxiliary library called **polarViz** to facilitate the visualization of the other packages, and a metapackage called **polArverse** that emulates the operation of tidyverse for loading the set of associated libraries.

**Meeting of universes: the polArverse and R-Universe**

During this process, I incorporated the R-Universe into the workflow. It provides an autonomous and self-managed space in which to publish these new packages (no longer simply as GitHub repositories) and provides support for testing on different operating systems and software versions. Undoubtedly, it is an important tool for better maintenance.

**polAr R-Universe**

**We continue to take important steps for our universe:**

First, we returned to CRAN! In February 2023, we shipped geoAr, which, as noted, provides access to and facilitates the processing of spatially relevant data. A version under development (accessible from the R-Universe) also allows you to interact with the API of the Argentine Geographic Data Standardization Service (Servicio de Normalización de Datos Geográficos de Argentina).
On the other hand, I added a pending tool to the polAr toolbox to work with census data: censAr. The package, currently under development, is based on the code from censo2017 developed by Mauricio “Pacha” Vargas and part of the rOpenSci suite.

**Summarizing my experience**

- I started working with this language - or dialect, as one of my colleagues likes to say regarding the antecedent of S from Bell in the 1970s - a little by chance.

- I quickly incorporated it as a work tool, in multiple facets: as an analyst, consultant and teacher.

- The R community - in particular all my experiences with LatinR - were a determining factor for my development. It certainly made the process much easier.

- I think it is possible to make improvements to the process; to support CRAN becoming an even better space for the R ecosystem and to make the R development experience more enjoyable and predictable.

- The R-Universe project to enhance the publishing and discovery experience of packages developed in R is invaluable.

The polarverse was a side project where I worked in my free time and without funding. It emerged almost playfully, as an experiment to learn how to develop software. We successfully published it in CRAN. We got frustrated and angry. We learned. We added new tools.

Now, with the support and encouragement provided by the great team at Fundar, in general, and the data area that I coordinate, in particular. We kept moving forward. Little by little. “Slowly, but steady”.