

CepalStatR: a package in R for access to ECLAC statistics

Henry Osorto^{1,2} 

¹Facultad de Postgrado, Universidad Tecnológica Centroamericana, UNITEC, Honduras, henry.osorto@unitec.edu.hn

²Carrera de Economía, Universidad Nacional Autónoma de Honduras, UNAH, Honduras, henry.osorto@unah.edu.hn

Abstract– The purpose of this study is to implement a new function package in R code for the access and analysis of data from the Economic Commission for Latin America and the Caribbean CEPALSTAT. The methodology consisted of making use of the Open Data/API and generating queries that allowed obtaining the descriptions of the thematic areas, metadata and databases of each one of the indicators, in addition functions were included that allow the creation of frequency tables, charts and maps. The results of this article include visualization of some indicators that can be used through the CepalStatR package. Through this new package, a new tool is provided that facilitates access to an important database, which in turn will contribute to the research processes for users of the statistical language R.


Keywords: CEPALSTAT, R language, Statistics, databases, ECLAC.

Digital Object Identifier: (only for full papers, inserted by LACCEI).

ISSN, ISBN: (to be inserted by LACCEI).

DO NOT REMOVE

CepalStatR: un paquete en R para el acceso a estadísticas de la CEPAL

Henry Osorto^{1,2} 

¹Facultad de Postgrado, Universidad Tecnológica Centroamericana, UNITEC, Honduras, henry.osorto@unitec.edu.hn

²Carrera de Economía, Universidad Nacional Autónoma de Honduras, UNAH, Honduras, henry.osorto@unah.edu.hn

Resumen— *El propósito de este estudio es implementar un nuevo paquete de funciones en código R para el acceso y análisis de datos de la Comisión Económica para América Latina y el Caribe CEPALSTAT. La metodología consistió en hacer uso de la Open Data/API y generar consultas que permitieron obtener las descripciones de las áreas temáticas, metadatos y bases de datos de cada uno de los indicadores, además se incluyeron funciones que permite la creación de tablas de frecuencias, gráficos y mapas. En los resultados de este artículo se incluyen visualización de algunos indicadores que podrán utilizarse por medio del paquete CepalStatR. Por medio de este nuevo paquete se brinda una nueva herramienta que facilite el acceso a una importante base de datos, que a su vez contribuirá con los procesos de investigación para los usuarios del lenguaje estadístico R.*

Palabras clave-- CEPALSTAT, Lenguaje R, Estadísticas, bases de datos, CEPAL.

I. INTRODUCCIÓN

Hacer uso de bases de datos, tanto de macrodatos como de microdatos, es una práctica muy común en el desarrollo de investigaciones científicas, de hecho, una buena base de datos es una parte importante de la investigación. Para el caso de los datos macroeconómicos, su uso se generalizó desde la década de 1970 principalmente para la construcción de modelos econométricos [1], [2]. Gracias a los avances que han existido en la creación de bases de datos por parte de instituciones nacionales y organizaciones internacionales que fomentan el desarrollo económico y social, hoy en día es posible acceder a números indicadores y variables de muchos países y en diferentes años. Esto ha permitido que, en lugar de trabajar con T observaciones de series de tiempo de N variables, donde frecuentemente T es grande y N es pequeño, ahora podemos analizar una gran cantidad de variables sin sacrificar información en la dimensión de series de tiempo [3]. A estos conjuntos de datos se les suele llamar entorno rico en datos, siempre que se satisfaga que T y N sean grandes [4]. Además de las ventajas de trabajar con la dimensión temporal y un mayor número de variables, los conjuntos de datos actuales puestos a disposición por organizaciones como Banco Mundial, BID, CEPAL, SEMCA, OIT entre otros, son capaces de realizar desagregaciones de diferentes variables, tal como lo suelen hacer los cortes transversales, por ejemplo, por países, poblaciones por sexo, edad, área urbana y rural etc. Este tipo de información suele ser llamada dimensión estructural. En tal sentido, muchos de los conjuntos de datos disponibles en los

sitios web de estas organizaciones ofrecen información que combina la dimensión temporal y estructural, es decir, datos de panel. Los beneficios de utilizar datos de panel incluyen la capacidad de capturar la heterogeneidad no observable, que puede estar conformada por efectos individuales específicos (diferencias entre países) y los efectos temporales [5].

Adicionalmente, muchas bases de datos macroeconómicas cuentan con una interfaz de programación de aplicaciones, conocida por sus siglas en inglés API. Las API se utilizan frecuentemente para ofrecer servicios empresariales a través de una variedad de dispositivos y canales digitales [6], permitiendo que los datos o servicios proporcionados por las empresas se comuniquen con diversas aplicaciones. Por ejemplo, al usar Google Maps se establece una conexión con una API que envía información sobre la ubicación de un lugar, así como las diferentes rutas y tiempos de llegada. Otras empresas, como Amazon, utilizan API para brindar información sobre sus productos. Esta tecnología permite a los desarrolladores crear aplicaciones innovadoras y eficientes que enriquecen la experiencia del usuario [7]. Sin embargo, el buen funcionamiento y diseño de estas aplicaciones dependen en gran medida de la arquitectura y documentación de las API, lo cual es crucial para una comprensión adecuada por parte de los desarrolladores [8].

Si bien para un desarrollador puede resultar relativamente sencillo y eficiente crear una aplicación que permita acceder a una API y obtener datos actualizados de manera remota, en el ámbito de la investigación científica, especialmente en lo que respecta al acceso a bases de datos compartidas a través de una API, el proceso puede resultar considerablemente más complejo para investigadores que carecen de conocimientos o experiencia en el desarrollo de aplicaciones

En lugar de eso, los investigadores a menudo se ven obligados a recurrir a descargas masivas y manuales de hojas de cálculo para utilizar los datos en sus estudios. Sin embargo, este enfoque puede presentar ciertas dificultades, ya que el procesamiento manual de grandes volúmenes de datos puede ser laborioso y propenso a errores. Además, este proceso en ocasiones requiere mucho tiempo y no es fácilmente automatizable. Probablemente bajo este mismo contexto se ha desarrollado el paquete `wbstats` [9], para el lenguaje de programación R. Este paquete brinda acceso a datos del Banco Mundial a través de su API, permitiendo a los investigadores trabajar con estos datos de manera más eficiente y conveniente, evitando la necesidad de descargar y procesar manualmente grandes conjuntos de datos.

Digital Object Identifier: (only for full papers, inserted by LACCEI).

ISSN, ISBN: (to be inserted by LACCEI).

DO NOT REMOVE

Siguiendo esta misma línea, el desarrollo del paquete CepalStatR en R responde a la necesidad de proporcionar a los investigadores un acceso más fácil y efectivo a los datos de la CEPAL a través de su API. Este paquete ofrece una serie de funciones diseñadas específicamente para simplificar el proceso de extracción y análisis de datos, permitiendo a los investigadores aprovechar al máximo la riqueza de información disponible en la base de datos de la CEPAL. Además de facilitar el acceso a los datos, esta herramienta fomenta una mayor transparencia y colaboración en la investigación científica al eliminar las barreras tradicionales en el acceso a la información.

Por tanto, el objetivo principal de este artículo es presentar el desarrollo y la funcionalidad de este paquete, destacando su potencial para impulsar los procesos de investigación científica que requieren el uso de datos provenientes de la CEPAL.

II. MATERIALES Y MÉTODOS

A. Desarrollo del paquete

Para desarrollar el paquete CepalStatR se hizo uso del lenguaje de programación R [11], mediante la interfaz de R Studio. Las funciones fueron creadas con la finalidad de establecer una conexión con la base de datos de la CEPAL, así como la elaboración de algunos análisis de tipo gráfico con variables e indicadores de dicha base. En el código desarrollado se emplearon funciones del código base de R, así como de paquetes orientados al procesamiento y visualización de información. En tal sentido, las consultas que realiza el paquete a la API de CEPALSTAT vienen dadas en formatos csv y json, por tanto, para este último, se hizo uso del paquete jsonlite [10]. Para realizar procesos de selección, filtrado uniones, resúmenes de datos se utilizaron los paquetes dplyr [12], tidyr [13], magrittr [14]. Para tareas en las que se requería procesar cadenas de texto se utilizaron funciones del paquete stringr [15], mientras que para la visualización html del conjunto de indicadores se recurrió al paquete gt [16]. En cuanto a las funciones orientadas a resultados gráficos, se usaron los paquetes ggplot2 [17] y gridExtra [18]. Los paquetes usados para la configuración y arquitectura de CepalStatR fueron usethis [19] y devtools [20]. Cada una de las funciones, así como los archivos auxiliares del paquete se alojaron en el repositorio GitHub: Henry-Osorto/CepalStatR. Mediante este repositorio es posible acceder al paquete y cada una de sus funciones.

B. Descripción del paquete

CepalStatR es un paquete que tiene como objetivo principal brindar acceso a la base de datos de CEPALSTAT mediante la API Pública de modo que se puedan realizar consultas de las áreas temáticas, metadata, notas, variables e indicadores por medio del lenguaje de programación R. Actualmente es un paquete en desarrollo que cuenta con cuatro funciones principales que brinda acceso a los datos: *viewer.indicators*, *call.indicators*, *countries*, *call.data* y dos funciones de creación de salida gráfica: *pyramids* y *ranking.sdg*.

En la figura 1, se muestra el proceso que se debe de seguir para poder hacer uso del paquete. Por tanto, una vez instalado, se recomienda hacer una inspección o exploración de los indicadores, luego elegir el tipo de consulta: consultar los datos para un objetivo determinado, o los indicadores mediante la salida gráfica.

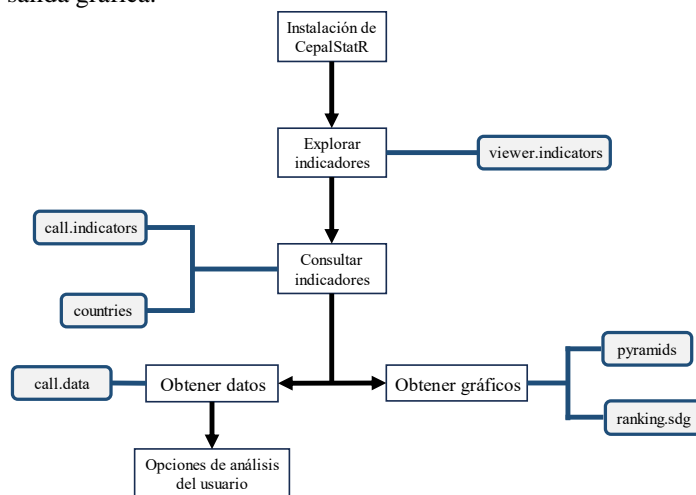


Fig. 1 Proceso de uso del paquete CepalStatR

III. RESULTADOS Y VISUALIZACIONES

A. Instalación del paquete

La descarga e instalación de este paquete CepalStatR se puede realizar mediante el repositorio GitHub a través del paquete devtools, ejecutando el siguiente código:

```
R> install.packages("devtools")
R> devtools::install_github("Henry-Osorto/CepalStatR")
R> library("CepalStatR")
```

B. Exploración de indicadores

Después de completar la instalación, se recomienda explorar los indicadores disponibles para identificar el nombre del indicador, el área temática a la que pertenecen, las dimensiones y subdimensiones asociadas, así como el número de indicador que será utilizado para desarrollar consultas a los datos. Para este propósito, se debe ejecutar la función *viewer.indicators*, que cuenta con un único parámetro lógico: *language.en*, el cual permite definir el idioma en el que CEPALSTAT proporciona los datos (inglés o español). Si se omite el parámetro o se establece como TRUE, se obtendrán datos en inglés; por otro lado, para seleccionar el idioma español se debe especificar FALSE.

```
R> viewer.indicators(language.en = FALSE)
```

El resultado de esta función será una tabla de salida que facilita la visualización de los indicadores y que puede abrirse en formato HTML para brindar mayor flexibilidad al usuario. En la figura 2 se proporciona una ilustración del resultado de la función.

Base de Datos de CEPALSTAT						
Área	Dimensión	Subdimensión	Indicador.1	Indicador.2	Indicador.3	Id del Indicador
Demográficos y sociales	Demográficos	Población	Población total, según sexo			1
Demográficos y sociales	Demográficos	Población	Población, según grupos de edad y sexo			31
Demográficos y sociales	Demográficos	Población	Relación de dependencia demográfica, según grupos dependientes y sexo			35
Demográficos y sociales	Demográficos	Población	Estructura de la población según sexo y grupos de edad			390

Fig. 2 Visualización de indicadores por medio de la función viewer.indicators

Por medio de esta tabla, es posible navegar y conocer la estructura de la base de datos en cuanto a la composición de las áreas, dimensiones y subdimensiones. En ocasiones, los indicadores tienen hasta cinco niveles de clasificación, comenzando desde la principal, que es el área. Sin embargo, cuando solo tienen tres niveles (como se muestra en la figura

2), el nombre del indicador aparecerá en la columna "Indicador.1". Análogamente, el nombre del indicador aparecerá en la siguiente columna si se cuenta con otro nivel de clasificación. Finalmente, en la columna "Id del Indicador", se podrá identificar el número de indicador que permitirá llamar los datos mediante las funciones posteriores.

Como complemento a las funciones de consulta de indicadores, el paquete cuenta con la función *call.indicators*, que permite obtener una tabla que puede ser almacenada como un data.frame y luego exportada a un formato CSV o XLSX, o bien realizar otras operaciones. Por ejemplo, el siguiente código muestra cómo se almacena la tabla en un objeto tipo data.frame para luego resumir la cantidad de indicadores, agrupados por área y dimensión.

```
R> indicadores <- call.indicators(language.en = F)
```

```
# Para resumir los datos por área y dimensión ejecute:
```

```
R> indicadores %>%
  group_by(Área, Dimensión) %>%
  count()
```

TABLA 1
NÚMERO DE INDICADORES POR ÁREA Y DIMENSIONES DE CEPALSTAT

Área	Dimensión	Indicadores
Ambientales	Asentamientos humanos y salud ambiental	4
	Calidad ambiental	1
	Cobertura terrestre, ecosistemas y biodiversidad	13
	Condiciones físicas	3
	Eventos naturales extremos y desastres	4
	Recursos biológicos	7
	Recursos energéticos	24
	Recursos hídricos	4
	Regulación y gobernanza ambiental	4
	Residuos	13
	Tierra	5
Demográficos y sociales	Demográficos	71
	Sociales	188
Económicos	Comercio Internacional y Balanza de Pagos	86
	Estadísticas sectoriales y productividad	25
	PIB y Cuentas Nacionales	87
	Precios	44
	Sector financiero y monetario	11
	Sector público	13
Temas transversales	Ciudades	50
	Cohesión social	18
	Estadísticas de la actividad agropecuaria, desarrollo rural y seguridad alimentaria y nutricional de Centroamérica y México	104
	Estadísticas de los Subsectores Eléctrico y de Hidrocarburos de Centroamérica y la República Dominicana	33
	Estadísticas subnacionales y de áreas pequeñas	129
	Género	89
	Objetivos de Desarrollo Sostenible (ODS)	701
	Pueblos indígenas y afrodescendientes	20
	Tecnologías de información y comunicación	79

Fuente: Elaboración propia con base en datos del CEPALSTAT.

Además de las funciones orientadas a la consulta de indicadores, se ha agregado la función `countries` con el propósito de devolver un `data.frame` con la lista de los países de los cuales se puede disponer de información en la mayoría de los indicadores. Esto ayudará a proporcionar el nombre de los países de manera adecuada en la aplicación de la función `pyramids` de este mismo paquete, o bien, para filtrar los países que se deseen seleccionar para algún propósito específico del usuario.

Es importante señalar que la base de datos CEPALSTAT realiza agregaciones para grupos de países en algunas variables o indicadores específicos. Dichas agrupaciones incluyen: América Latina y el Caribe, América Latina, El Caribe, América Central y América del Sur. Para acceder a la lista de países y agrupaciones, se puede ejecutar el siguiente código:

```
R> países <- countries(language.en = F)
R> View(países)
```

C. Consultar los datos – marcos de datos

Una vez que se conocen los indicadores y sus respectivos ID, se puede proceder a consultar los datos o valores del indicador mediante la función `call.data`. El parámetro principal de la función es `id.indicator`, al cual se le debe especificar el número del ID del indicador o variable que se desea consultar. Adicionalmente la función incluye el parámetro lógico `notes`, que ayuda a incorporar las notas metodológicas que puede tener un indicador. Por defecto las notas no se mostrarán a menos que se especifique `TRUE` si así se desea. Así mismo se cuenta con el parámetro `language.en` para la selección del idioma.

En el siguiente fragmento de código se brinda un ejemplo cómo consultar los datos del indicador población total según sexo, donde su ID es el número 1. Con dicho indicador se puede contar con las proyecciones de la División de Población de la CEPAL (CELADE) y la División de Población de las Naciones Unidas, las cuales cubren un horizonte temporal de 150 años, divididas en: estimaciones anteriores (1950-2015) y proyecciones futuras (2015-2100) [21].

```
R> df <- call.data(id.indicator = 1, language.en = F)
R> View(df)
```

Tal como se define en el proceso de uso del paquete `CepalStatR`, luego de hacer la consulta a los datos el usuario es libre de aplicar cualquier clase de análisis cuantitativo, tales como la elaboración de tablas resumen estadístico, cálculo de nuevos indicadores, análisis gráfico o ajustar regresiones con datos de panel cuando se consulten otras variables y establezcan las uniones de los conjuntos de datos. Para efectos ilustrativos, el siguiente fragmento de código muestra la creación de un

gráfico de líneas para los valores de la población por sexo desde el año 1950 al 2100. Antes de elaborar el gráfico, los datos fueron filtrados en la variable País, lo cual permitió dividir el gráfico en múltiples paneles o facetas mediante la función `facet_wrap`.

```
# Ejemplo de creación de gráfico de líneas:
R> library(tidyverse)
R> df %>%
  filter(País %in% c('América Latina y el Caribe',
                    'Honduras', 'Guatemala',
                    'El Salvador', 'Costa Rica',
                    'Nicaragua')) %>%
  ggplot(aes(y = value, x = Años, color = Sexo)) +
  geom_line() +
  facet_wrap(~ País, scales = 'free_y') +
  labs(y = 'Miles de Personas', x = 'Años', color = '') +
  theme(legend.position = 'top')
```

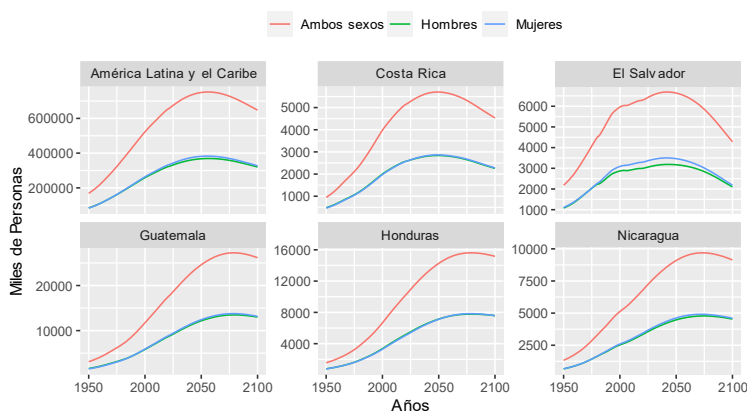


Fig. 3 Población total según sexo: Proyecciones al año 2,100

D. Consultar los datos – salidas gráficas

Otra forma de consulta a los datos es mediante el uso de las salidas gráficas. En tal sentido, la función `pyramids` proporciona la consulta al indicador población según grupos de edad y sexo, con la finalidad de elaborar pirámides poblacionales. Dicha función opera principalmente con dos parámetros `country` y `years`, mientras que seis restantes son opcionales. El parámetro `country` requiere de un cadena de texto que indique el nombre de un país al cual se le aplicará la función, por lo que es recomendable ejecutar `countries` previamente.

Así mismo, es necesario indicar al menos un año al que harán referencia los datos por medio del gráfico. A este respecto, cabe señalar que el parámetro `years` solicita que se proporcione un valor o vector numérico que pertenezca al conjunto $x = \{1, 2, 3, \dots, 31\}$, que no es más que la posición de un vector en el conjunto $y = \{1950, 1955, 1960, \dots, 2100\}$, el

cual hace referencia a los años (en intervalos de 5 años) que pone a disposición la base de datos CEPALSTAT. En tal sentido solo debe ser incluido uno o varios números que $\in x$. Para una mejor referencia de los años que se pueden consultar, se recomienda crear un vector de posición y otro de años, tal como se muestra en el siguiente fragmento de código.

```
R> posición <- seq(1, 31, 1)
R> años <- seq(1950, 2100, 5)
R> years <- cbind(posición, años)
R> View(years)
```

De igual manera se incorpora el parámetro `language.en`, para visualizar las etiquetas, leyendas y subtítulos en inglés o español. También se cuenta con un el parámetro `color`, con el cual se pueden modificar los colores que vienen por defecto en el gráfico, de modo que se debe especificar un vector de dos caracteres con los nombres de los colores que se desean seleccionar. Por defecto los colores incorporados en la función se especifican en escala RGB: `c('#B20B27', '#0A1873')`. Sin

embargo, puede modificar los colores utilizando los nombres convencionales (`red`, `yellow`, `blue`, `green`, etc.).

Se incorporó el parámetro `save`, si el usuario desea guardar el gráfico en formato de imagen PNG, el cual se alojará en el directorio `setwd('~')` con el nombre 'Pirámide Poblacional' si se eligió el idioma español o 'Population pyramids' si se omitió `language.en` o se seleccionó `TRUE`. Como complemento del parámetro `save`, se incluyeron los parámetros `height` y `width` para modificar las dimensiones altura y anchura del gráfico, si se decidió guardar dicho gráfico.

En el siguiente fragmento de código se brinda un ejemplo de la consulta de pirámides poblacionales de Honduras en los años 1950, 1970, 1995, 2020, 2050 y 2100. Aprovechando la opción de guardado que posee la función, se generó una imagen en formato PNG de las pirámides agrupadas, las cuales se ilustran en la figura 4.

```
R> pyramids(country = 'Honduras',
            years = c(1, 5, 10, 15, 21, 31),
            save = T, height = 10, width = 15)
```

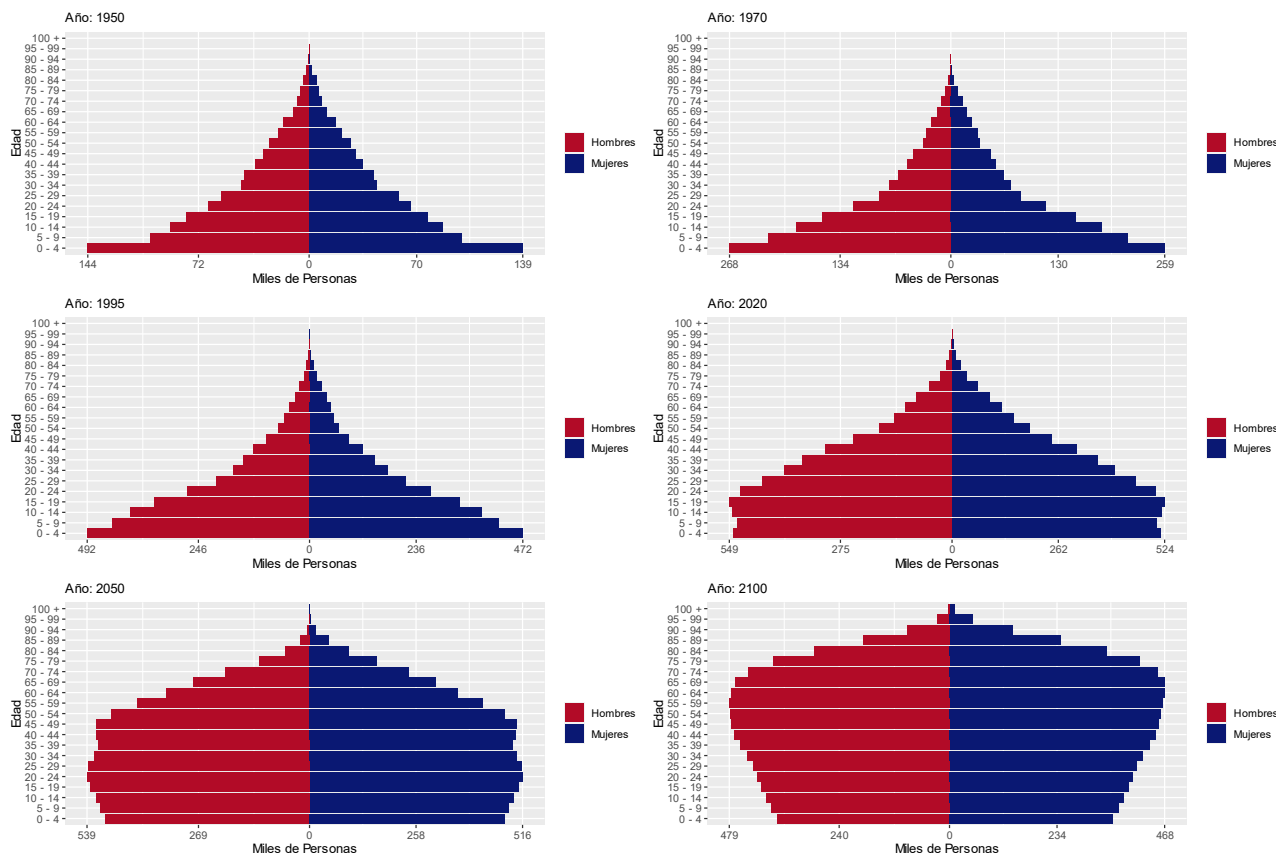


Fig. 4 Pirámides poblacionales de Honduras

Por su parte a función *ranking.sdg* permite consultar de forma gráfica los indicadores que pertenecen al área 'Temas Transversales' en la dimensión de 'Objetivos de Desarrollo Sostenible (ODS)', donde su finalidad es mostrar el logro alcanzado en el indicador comparado por países y los agregados de países siempre que estén disponibles. En tal sentido, la función sigue un proceso en el cual, una vez que se determina el indicador ODS, se procede a clasificar si el valor del indicador es un porcentaje o no, esto con el fin de segmentar o filtrar las agrupaciones de la variable país. Posteriormente la función elaborará el gráfico con la opción de que sea guardado. En la figura 5 se ilustra el proceso que realiza la función *ranking.sdg*.

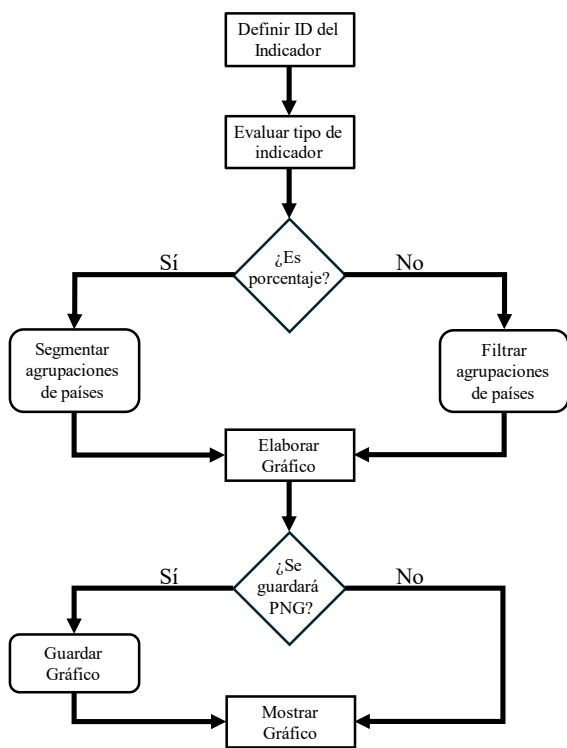


Fig. 5 Procesamiento de la función *ranking.sdg*

El principal parámetro que se requiere para esta función es *id.indicator*, el cual admitirá únicamente los ID de indicadores de la dimensión de Objetivos de Desarrollo Sostenible (ODS). El resto de los parámetros son opcionales de acuerdo con las necesidades del usuario. En ese sentido, está incorporado en la función el parámetro *language.en* para determinar el idioma de las etiquetas, leyendas, títulos y notas. El parámetro *title*, es un valor lógico que permite activar o desactivar la visualización del título del gráfico. De forma predeterminada, la opción de visualización del título es *TRUE*. Como complemento de *title*, se cuenta con el parámetro *size.title* para ajustar el tamaño del título del gráfico. También la función dispone del parámetro lógico *caption*, que permite activar o desactivar la visualización de una nota metodológica que se ha elaborado para indicar el último año en el que se encuentra disponible el valor del indicador. Por defecto la opción para visualizar la nota es *TRUE*.

En cuanto al aspecto visual se diseñó el parámetro *color*, que sirve para cambiar el color de las barras del gráfico para todos los países, mientras que *color.gc* ayuda a cambiar el color de los grupos de países, siempre y cuando se encuentren disponibles para el indicador y su valor sea un porcentaje. De igual manera se incluyó en esta función el parámetro lógico *save*, brindando la opción de guardar el gráfico en formato de imagen PNG, el cual se alojará en el directorio *setwd('~/.')* con el nombre 'Logro del indicador 'ID' si se eligió el idioma español, o con el nombre 'Indicator achievement 'ID' si se omitió *language.en* o se seleccionó *TRUE*. Así mismo se dispone de los parámetros *height* y *width* para modificar las dimensiones altura y anchura del gráfico, si se decidió guardar dicho gráfico. En el siguiente fragmento de código se brinda un ejemplo del grafico que se obtiene por medio de la función *ranking.sdg*, mismo que se ilustra en la figura 6.

```
R> ranking.sdg(id.indicator = 4028, language.en = F,
               save = T)
```

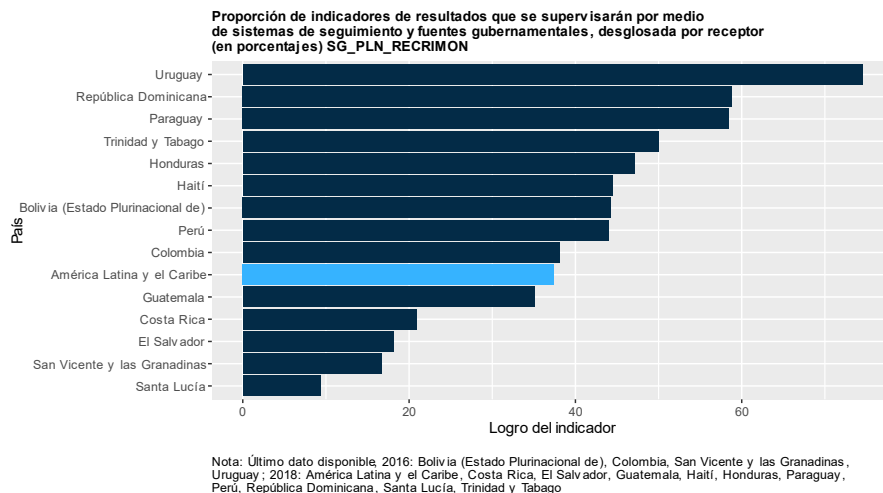


Fig. 6 Salida de la función *ranking.sdg*

IV. DISCUSIÓN Y CONCLUSIONES

La base de datos proporcionada por la CEPAL constituye una herramienta valiosa para investigadores, académicos y responsables de la toma de decisiones. La mejora en la disponibilidad y accesibilidad de los datos a través del paquete CepalStatR tiene el potencial de impulsar avances significativos tanto en la investigación académica como en el diseño e implementación de políticas públicas en la región de América Latina y el Caribe. Los investigadores podrán realizar análisis más detallados y contextualizados, abordando preguntas complejas y multidimensionales, así como explorando relaciones causales sobre una amplia gama de temas socioeconómicos, que incluyen el crecimiento económico, la pobreza, la desigualdad, la educación y la salud. En el ámbito de las políticas públicas, el acceso mejorado a los datos de la CEPAL a través de CepalStatR proporcionará a los responsables de la toma de decisiones información crítica para la formulación, monitoreo y evaluación de políticas orientadas al desarrollo sostenible y la reducción de la pobreza en la región. La actualización y desglose de los datos a nivel nacional y regional permitirá una comprensión más profunda de los desafíos específicos que enfrentan los países de América Latina y el Caribe, así como la identificación de áreas prioritarias de intervención. Además, la capacidad de realizar análisis comparativos entre países y a lo largo del tiempo facilitará la identificación de buenas prácticas y lecciones aprendidas que pueden informar políticas efectivas y basadas en evidencia. En este contexto, el paquete CepalStatR se convierte en una herramienta poderosa que puede contribuir significativamente a la generación de conocimiento y la toma de decisiones informada en la región de América Latina y el Caribe. Al facilitar el acceso y la utilización de los datos de la CEPAL, este paquete tiene el potencial de impulsar el desarrollo económico y social en la región, así como de contribuir al logro de los Objetivos de Desarrollo Sostenible a nivel nacional y regional.

El proyecto CepalStatR representa un primer paso importante en el establecimiento de conexiones entre la base de datos CEPALSTAT y lenguaje estadístico R, lo cual ofrece una serie de ventajas significativas debido a la capacidad de procesamiento y versatilidad de este lenguaje de programación. Aunque el paquete está diseñado específicamente para facilitar el acceso a los datos de la CEPAL a través de su API, su desarrollo continuará avanzando. Entre las futuras mejoras previstas se encuentra la incorporación de funciones del paquete `panelView` [22], que proporciona soluciones simples y prácticas para visualizar datos de panel. También se contempla la implementación de funciones para crear mapas utilizando los paquetes `tmap` [23] y `sf` [24]. Además, se explorarán funciones relacionadas con técnicas de regresión para ampliar aún más las capacidades analíticas del paquete.

En relación con las funciones existentes, se buscará mejorar la función `call.data` para permitir la conexión con múltiples indicadores simultáneamente. Esto evitará la necesidad de ejecutar múltiples consultas y realizar uniones de bases de datos, lo que mejorará la eficiencia y la facilidad de

uso del paquete. Asimismo, se pretende potenciar la función `pyramids` para facilitar la comparación al permitir la generación de pirámides poblacionales para varios países en un mismo año. Estas mejoras y ampliaciones previstas permitirán que el paquete CepalStatR continúe siendo una herramienta útil y relevante para la comunidad de investigadores y analistas que trabajan con datos de la CEPAL en el entorno de R.

REFERENCES

- [1] J. Galí and M. Gertler, "Macroeconomic Modeling for Monetary Policy Evaluation," *Journal of Economic Perspectives*, vol. 21, no. 4, pp. 25–45, 2007, doi: 10.1257/jep.21.4.25.
- [2] C. O. Renfro, "Economic database systems: further reflections on the state of the art 1," *JEM*, vol. 23, no. 1, pp. 43–85, 1997, doi: 10.3233/JEM-1997-23103.
- [3] M. W. McCracken and S. Ng, "FRED-MD: A Monthly Database for Macroeconomic Research," *Journal of Business & Economic Statistics*, vol. 34, no. 4, pp. 574–589, 2016, doi: 10.1080/07350015.2015.1086655.
- [4] B. S. Bernanke and J. Boivin, "Monetary policy in a data-rich environment," *Journal of Monetary Economics*, vol. 50, no. 3, pp. 525–546, 2003, doi: 10.1016/S0304-3932(03)00024-2.
- [5] M. Mayorga M. and E. Muñoz, "La técnica de datos de panel una guía para su uso e interpretación," Banco Central de Costa Rica, 2000.
- [6] B. De, "API Management," in *API Management*, B. De, Ed., Berkeley, CA: Apress, 2023, pp. 27–47.
- [7] B. De, "Introduction to APIs," in *API Management*, B. De, Ed., Berkeley, CA: Apress, 2023, pp. 1–26.
- [8] B. De, "API Documentation," in *API Management*, B. De, Ed., Berkeley, CA: Apress, 2023, pp. 93–125.
- [9] J. Piburn, *wbstats*: Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 2016.
- [10] J. Ooms, "The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects," 2014. [Online]. Available: <http://arxiv.org/pdf/1403.2805.pdf>
- [11] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria. [Online]. Available: <https://www.r-project.org/>
- [12] Hadley Wickham, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan, *dplyr: A Grammar of Data Manipulation*. [Online]. Available: <https://cran.r-project.org/package=dplyr>
- [13] Hadley Wickham, Davis Vaughan, and Maximilian Girlich, *tidyr: Tidy Messy Data*. [Online]. Available: <https://cran.r-project.org/package=tidyr>
- [14] Stefan Milton Bache and Hadley Wickham, *magrittr: A Forward-Pipe Operator for R*. [Online]. Available: <https://cran.r-project.org/package=magrittr>
- [15] Hadley Wickham, *stringr: Simple, Consistent Wrappers for Common String Operations*. [Online]. Available: <https://cran.r-project.org/package=stringr>

- [16] Richard Iannone, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, and JooYoung Seo, *gt: Easily Create Presentation-Ready Display Tables*. [Online]. Available: <https://cran.r-project.org/package=gt>
- [17] Hadley Wickham, *ggplot2: Elegant Graphics for Data Analysis*: Springer-Verlag New York. [Online]. Available: <https://ggplot2.tidyverse.org/>
- [18] Baptiste Auguie, *gridExtra: Miscellaneous Functions for "Grid" Graphics*. [Online]. Available: <https://cran.r-project.org/package=gridExtra>
- [19] Hadley Wickham, Jennifer Bryan, Malcolm Barrett, and Andy Teucher, *usethis: Automate Package and Project Setup*. [Online]. Available: <https://cran.r-project.org/package=usethis>
- [20] Hadley Wickham, Jim Hester, Winston Chang, and Jennifer Bryan, *devtools: Tools to Make Developing R Packages Easier*. [Online]. Available: <https://cran.r-project.org/package=devtools>
- [21] Comisión Económica para América Latina y el Caribe (CEPAL) and Naciones Unidas, *CEPALSTAT Bases de Datos y Publicaciones Estadísticas: Ficha técnica. Estadísticas e Indicadores Sociales. Población total, según sexo*. [Online]. Available: https://statistics.cepal.org/portal/cepalstat/technical-sheet.html?lang=es&indicator_id=1
- [22] H. Mou, L. Liu, and Y. Xu, "Panel Data Visualization in R (`panelView`) and Stata (`panelview`)," *J. Stat. Soft.*, vol. 107, no. 7, 2023, doi: 10.18637/jss.v107.i07.
- [23] M. Tennekes, "tmap : Thematic Maps in R," *J. Stat. Soft.*, vol. 84, no. 6, 2018, doi: 10.18637/jss.v084.i06.
- [24] E. Pebesma, "Simple Features for R: Standardized Support for Spatial Vector Data," *The R Journal*, vol. 10, no. 1, p. 439, 2018, doi: 10.32614/RJ-2018-009.