



# *importinegi:*

un paquete de *R* para descargar y gestionar bases de datos del INEGI

## *importinegi:*

an *R* Package to Download and Manage INEGI's Data Sets

César Rentería\*

\* Centro de Investigación y Docencia Económicas, [crenteria@albany.edu](mailto:crenteria@albany.edu)

Este artículo presenta *importinegi*, paquete que contribuye a facilitar la búsqueda, descarga y gestión de los datos abiertos del Instituto Nacional de Estadística y Geografía. Se describen todas las funciones y se acompañan con ejemplos. Fue desarrollado en plataforma abierta y busca apoyar al creciente ecosistema de ciencia abierta para la investigación.

**Palabras clave:** ciencia abierta; datos abiertos; bases de datos INEGI; *R*.

Recibido: 8 de agosto de 2019.

Aceptado: 3 de junio de 2020.

## 1. Introducción

El Instituto Nacional de Estadística y Geografía (INEGI) produce y provee un vasto conjunto de datos abiertos a partir de diversos proyectos. Aunque ha realizado un esfuerzo importante durante los últimos años para mejorar su disponibilidad y usabilidad, la integración de estos en ambiente de programación de *software* estadístico es todavía un área en desarrollo. Para ello, se cuenta con *R*, que es un *software* de licencia abierta que ha crecido considerablemente en popularidad y uso (especialmente en ciencias sociales).

Este artículo presenta el paquete de *R importinegi*, cuyo objetivo es facilitar la búsqueda, descarga y gestión de bases de datos abiertos del INEGI, en particular para su uso en la investigación y la enseñanza. Asimismo, al proveer funciones de código abierto —y en un *software* libre—, contribuye a fortalecer la adopción de mejores prácticas de *reproducibilidad* de la ciencia (Peng, 2011).

Facilitar el acceso y análisis de estos a través de paquetes estadísticos es un esfuerzo con muchos beneficios (Boettiger, Chamberlain *et al.*, 2015; Gandrud, 2016); por ejemplo: contribuye a expandir la *reproducibilidad* y transparencia de un artículo de investigación desde los datos en bruto hasta la publicación final; al simplificar, estandarizar y automatizar los procesos de gestión, también coadyuva en el logro eficiente del trabajo de investigación; asimismo, promueve la compartición

This article presents *importinegi*, a package that helps to facilitate the search, download and management of the *National Institute of Statistics and Geography's* open data. All functions are described and illustrated with examples. It was developed on an open platform and seeks to support the growing ecosystem of open science for research.

**Key words:** open science; open data; INEGI databases; *R*.

de código y la colaboración académica, minimiza errores humanos en el emparejamiento de múltiples bases de datos relacionadas y reduce problemas de acceso por cuestiones de compatibilidad o licenciamiento de formatos de archivo.

Diversos organismos a nivel internacional han desarrollado, ya sea institucionalmente o a través de sus usuarios, paquetes de *R* para facilitar el acceso a sus bases de datos y para promover los principios de *reproducibilidad* de la ciencia, por ejemplo: el Banco Mundial (Arel-Bundock, 2013), la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, por sus siglas en inglés) (Gheri & Kao, 2014), la oficina Eurostat de la Unión Europea (Lahti *et al.*, 2017) o el Buró del Censo en Estados Unidos de América (Walker, 2018).

Con *importinegi* se busca contribuir con la *reproducibilidad* de la ciencia y ampliar la base de usuarios de los datos abiertos del INEGI. Este trabajo provee una visión general de las características del paquete y describe cada una de las funciones de la versión 1.0.0.

## 2. El paquete *importinegi*

Se creó para facilitar la descarga y uso de las bases de datos públicas del INEGI. Entre otras ventajas, permite integrar múltiples de estas con un identificador único en común, consolidar los datos en diferentes niveles de agregación (por ejemplo, vi-

vienda, localidad, municipio o entidad federativa) y automatizar la codificación de valores perdidos en el ambiente de R.

La versión 1.0.0 de *importinegi* se publicó en el *Comprehensive R Archive Network (CRAN)* el 5 de agosto de 2019. Permite descargar y gestionar las bases de datos de los siguientes proyectos estadísticos del INEGI:

- Censo (y Conteo) de Población y Vivienda.
- Censo Nacional de Gobiernos Municipales y Delegacionales (CNGMD).<sup>1</sup>
- Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH).
- Encuesta Nacional de Ocupación y Empleo (ENOE).

Asimismo, hace posible el acceso a las siguientes fuentes de información georreferenciada:

- Marco Geoestadístico Nacional (MGN).
- Red Nacional de Caminos (RNC).

La versión 1.0.0 *importinegi* puede descargarse desde el *CRAN* tecleando la función de instalación `install.packages()`. En el siguiente bloque de código se muestra cómo hacerlo:

```
install.packages("importinegi")
library(importinegi)
```

El paquete continúa en constante desarrollo, incorporando más datos abiertos para su descarga y refinando las funciones vigentes. La versión beta con las últimas actualizaciones se puede descargar directamente del repositorio de *Github* *crenteriam/importinegi*. En el siguiente bloque de código se muestra cómo instalarlo desde *Github*:

```
install.packages("devtools")
devtools::install_github(
  "crenteriam/importinegi")
library(importinegi)
```

<sup>1</sup> Ahora denominado Censo Nacional de Gobiernos Municipales y Demarcaciones Territoriales de la Ciudad de México.

La versión 1.0.0 cuenta con una extensa documentación de cada función y los valores, a la cual se puede acceder con la función de ayuda seguida del nombre de esta (por ejemplo, *importinegi*). También, es posible tener acceso a la viñeta del paquete mediante `browseVignettes("importinegi")`.

Cada función en el paquete corresponde a una fuente de datos distinta, por ejemplo, `enoe()` descarga y gestiona información de la ENOE.

En las siguientes secciones se describirá la sintaxis y uso de las funciones disponibles en la versión 1.0.0 de *importinegi*.

### 3. Sintaxis

El nombre de cada función representa las palabras clave o siglas del proyecto INEGI (como `censo_municipal()` o `enoe()`). En ocasiones, a esta nomenclatura le sigue una característica específica de este, por ejemplo: se puede acceder al Censo de Población y Vivienda a través de cinco formatos de microdatos, siendo ITER uno de ellos; por lo tanto, las funciones del Censo tienen una palabra clave adicional indicando el formato de microdatos: para llegar al de ITER es `censo_poblacion_iter()`.

En el caso de la información georreferenciada, el nombre de la función comienza con las siglas SIG (de sistema de información geográfica). Estas son sucedidas por una palabra clave del proyecto geoestadístico, por ejemplo, para acceder a los datos de la Red Nacional de Caminos es `sig_caminos()`.

Existe una función especial, `catalogo_inegi()`, que permite explorar el catálogo de proyectos estadísticos del INEGI, así como la documentación completa de cada base de datos. La función accede a la Red Nacional de Metadatos (RNM), que es una plataforma de difusión y consulta de cada proyecto. Su sintaxis es sencilla. El único parámetro requerido es el número del proyecto (id) a consultar. En el siguiente bloque de código se ac-

cede (a través del navegador por *default* del sistema operativo) a la documentación del *Censo Nacional de Derechos Humanos Estatal 2018* mediante su id:

```
# Accede a la documentacion del
Censo Nacional de
# Derechos Humanos Estatal 2018
catalogo_inegi(id = 443)
```

Tecleando la función `catalogo_inegi()` (sin incluir nada en el paréntesis) se puede descargar y almacenar en un objeto de *R* una base de datos con el catálogo de proyectos del INEGI. Esta provee el id, su nombre, los años en que se comenzó y finalizó el levantamiento, la fecha de creación de los metadatos en la RNM y la fecha de última actualización de estos. El siguiente bloque de código ilustra la descarga de la base de datos del catálogo de proyectos y su almacenamiento en un objeto de *R*:

```
# Descarga y almacena el catálogo
de proyectos
# estadisticos del INEGI
catalogoINEGI = catalogo_inegi()
```

### 3.1 Censo (y Censo) de Población y Vivienda

Estos operativos proporcionan información sobre individuos, viviendas y características de las localidades en México. El INEGI ofrece varias formas de acceso a los microdatos de los censos y conteos. El paquete *importinegi* tiene una función por cada

una de estas modalidades de acceso. En el cuadro 1 se presentan los tipos de microdatos disponibles y su función correspondiente.

Todas las funciones para los cinco tipos de microdatos de los censos tienen la misma sintaxis. Aunque estas varían en los parámetros aplicables, todas comparten dos indispensables.

El primero es el año (*year*) en formato numérico. Los datos que se tienen comienzan a partir de 1990 (e incluyen los conteos), pero no todos los tipos de microdatos están presentes para todos los años censales. La documentación de cada función provee los años disponibles para cada tipo de microdatos del censo.

El segundo es la entidad federativa (*estado*) en formato alfanumérico. La selección por *default* es *Nacional*, que provee los datos para todos los estados. Estos se deben anotar con letra capital y las entidades con múltiples palabras se escriben espaciadas y sin acentos (por ejemplo, San Luis Potosí). Para el caso de la Ciudad de México (anteriormente Distrito Federal) es CDMX.

El resto de los parámetros varía entre funciones. El cuadro 2 presenta todos los disponibles en cada función. El primer grupo de ellos (*total estado*, *total municipio*, *total localidad* y *total-ageb*) requiere información lógica (*TRUE* o *FALSE*), donde *FALSE* es el *default* y significa que la base de datos excluye los resultados agregados a nivel estado, municipio, localidad o AGEB, respectivamente. El segundo grupo (localidades y manzana) también necesita información

Cuadro 1

#### Funciones disponibles para los censos y conteos de población y vivienda

Función	Tipo de microdatos
<code>censo_poblacion_iter()</code>	Principales resultados por localidad (ITER).
<code>censo_poblacion_muestra()</code>	Muestra del cuestionario ampliado (muestra).
<code>censo_poblacion_ageb()</code>	Principales resultados por Área Geoestadística Básica (AGEB).
<code>censo_poblacion_urbano()</code>	Resultados sobre infraestructura y características del entorno urbano (urbano).
<code>censo_poblacion_rural()</code>	Resultados sobre localidades con menos de 5 mil habitantes (rural).

### Parámetros disponibles por tipo de microdatos del Censo

Tipo de microdatos	Parámetros
ITER	año, estado, total estado, total municipio, localidades
Muestra	año, estado, muestra
AGEB	año, estado, total estado, total municipio, total localidad, total ageb, manzana
Urbano	año, estado
Rural	año, estado

lógica, donde TRUE es el *default* y significa que la base de datos incluye información a nivel manzana o localidad, en ese orden.

El objetivo de los parámetros lógicos es ofrecer flexibilidad al usuario en la descarga. Al manipularlos, este puede limitar la descarga de datos a solo uno de los niveles de gobierno o cualquier combinación de estos. En el siguiente bloque de código se presentan cuatro ejemplos de descarga entre diversos niveles de gobierno:

```
# Default: a nivel localidad, pero sin datos a nivel estado
# ni municipio
censo.jalisco = censo_poblacion_iter
(year = 2010, estado = "Jalisco")
```

```
# Datos a nivel estado, municipio y localidad
censo.jalisco = censo_poblacion_iter
(year = 2010,
estado =
"Jalisco", totalestado = TRUE,
totalmunicipio = TRUE)
```

```
# Datos a nivel estado y municipio, pero sin datos a
# nivel localidad
censo.jalisco = censo_poblacion_iter
(year = 2010, estado =
```

```
"Jalisco", total estado = TRUE,
total municipio = TRUE,
localidades = FALSE)
```

```
# Datos a nivel nacional
censo.jalisco = censo_poblacion_iter
(year = 2010,
estado = "Nacional ")
```

En la muestra del Censo, además del año y el estado, un tercer parámetro indispensable es **muestra**, que representa la unidad de análisis, las cuales pueden ser **Migrantes**, **Personas**, **Viviendas** u **Hogar**. En el siguiente bloque de código se ilustra la descarga del cuestionario de personas para el Censo 2010 en Jalisco:

```
# Descarga cuestionario de personas
muestra.jalisco = censo_poblacion_
muestra (year = 2010,
estado = "Jalisco", muestra =
"Personas")
```

### 3.2 Censo Nacional de Gobiernos Municipales y Delegacionales

Es un proyecto estadístico sobre la gestión y desempeño de las entidades gubernamentales mexicanas a nivel municipal y delegacional. El CNGMD cubre cuatro temáticas: ayuntamiento, administra-

ción pública municipal o delegacional, seguridad y justicia. Cada una representa una fuente de información independiente con múltiples bases de datos, por ejemplo, para su edición 2011, la del ayuntamiento contiene dos conjuntos de datos con variadas bases: comisiones e integrantes.

La función `censo_municipal()` descarga las bases de datos de las cuatro temáticas y tiene tres parámetros indispensables. Primero, es necesario establecer en formato numérico el año del levantamiento del CNGMD, que puede ser 2011, 2013, 2015 o 2017. En el segundo, `fuentes`, se tiene que definir en formato alfanumérico la temática o fuente de la base de datos requerida. Por último, en el parámetro `datos` se debe definir en formato alfanumérico el conjunto de bases de datos.

La relación entre los datos, conjunto de datos y fuentes puede ser complicada, en especial a partir del CNGDM 2015. Como referencia, el cuadro 3 presenta dicha relación para el CNGDM 2011.

Se recomienda consultar la disponibilidad de datos por cada fuente, conjunto de datos y año de levantamiento directamente en la documentación en línea del INEGI. Se puede acceder a esta al teclear la función `censo_municipal()` (sin incluir nada entre los paréntesis), como se muestra en el siguiente bloque de código:

```
# Acceder a la documentación en
# línea del
# CNGDM censo_municipal()
```

Al elegir una fuente y conjunto de datos específicos, la función descarga una lista de *R* con el conjunto de bases de datos correspondiente. En el siguiente bloque de código se ilustra la descarga del conjunto de datos *comisión*, de la fuente *ayuntamiento*. El resultado de la descarga es una lista de *R* con cinco bases de datos (`ACT_AYUN`, `COM_AYUN`, `COM_TASU`, `IN_TEASU`, `INI_AYUN`):

```
# Descargar conjunto de datos
# "comisión" del
# censo 2011 datos.comisión =
# censo_municipal(year = 2011,
# fuentes = "ayuntamiento", datos =
# "comisión")
```

Esta lista, sin embargo, no es útil para el análisis estadístico; cada base de datos debe extraerse de la lista y almacenarse en un objeto de *R* distinto. Esto se logra, primero, identificando la posición de la base de datos en la lista con la función `summary()`. Posteriormente, se debe extraer la base de datos utilizando su posición dentro de la lista, como se ilustra en el siguiente bloque de código:

```
# Explorar la posición de cada
# base de datos en la
# lista summary(datos.comisión)

# Extraer la base de datos
# ACT_AYUN, que está
# en la posición número uno
# datos.actas = datos.comisión[[1]]
```

Cuadro 3

### Relación entre fuentes y conjuntos de datos en el CNGMD del 2011

Fuentes	Datos
Ayuntamiento	Comisión, Estructura
Administración	Estructura, Funciones, Marco, Participación, Recursos, Trámites, Transparencia
Seguridad	Ejercicio, Infraestructura, Recursos
Justicia	Ejercicio, Infraestructura, Recursos

### 3.3 Encuesta Nacional de Ingresos y Gastos de los Hogares

La ENIGH proporciona datos estadísticos sobre las entradas y salidas económicas de las unidades de observación en cuanto a monto, procedencia y distribución. Adicionalmente, provee información acerca de las características sociodemográficas de los integrantes del hogar. En la actualidad, existen tres variaciones de la Encuesta: 1) Tradicional (1984-2014), 2) Nueva Construcción (2008-2014) y 3) Nueva Serie (2016). El paquete *importinegi* 1.0.0 provee acceso únicamente a la versión Nueva Construcción.

La función `enigh_nuevaconstruccion()` descarga las bases de datos de esta. La ENIGH cuenta con diversas en tres niveles de unidad de análisis: vivienda, hogar y personas. Como se muestra en la figura 1, la Encuesta se compone de varias cadenas de bases de datos subordinadas, por ejemplo, la de trabajos tiene dos de estas (*agro* y *no agro*) y la de *hogares*, tres (*gasto hogar*, *erogaciones* y *gasto tarjeta*). Todas contienen información que puede ser integrable en una sola matriz de datos, aunque esta labor es compleja. La principal ventaja de esta función es que reduce dicha complejidad al descargar e integrar todas estas bases de datos en una sola. La función las une a todas siguiendo el diagrama de relaciones en la figura 1.

La función requiere el parámetro del año del levantamiento de la ENIGH, que puede ser 2008, 2010, 2012 o 2014. Descarga cada una de las bases de datos que componen la Encuesta. La función tiene dos parámetros indispensables. El primero es *año del levantamiento*. El segundo es *datos*, que define la base de datos a descargar. Las opciones son *viviendas*, *hogares*, *concentrado*, *erogaciones*, *gasto hogar*, *gasto tarjetas*, *poblacion*, *ingresos*, *gasto persona*, *trabajos*, *agro* y *no agro*. El siguiente bloque de código provee varios ejemplos del uso de esta función para descargar datos de la ENIGH 2014, Nueva Construcción:

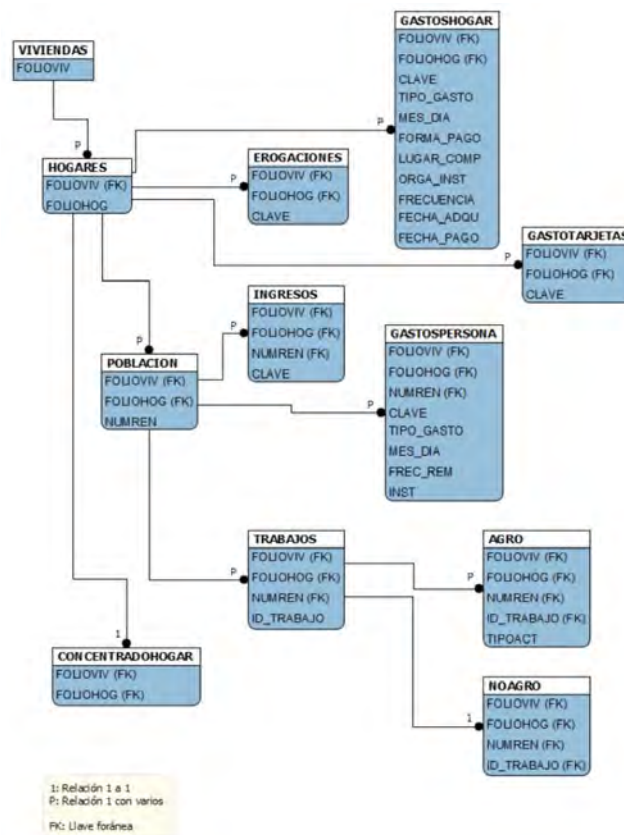
```
# Default: descargar datos de viviendas
```

```
viviendas14 = enigh_nuevaconstruccion(
year = 2014, datos = "viviendas")
```

```
# Descargar datos de hogares
hogares14 = enigh_nuevaconstruccion(
year = 2014, datos = "hogares")
```

```
# Descargar datos de agro
agro14 = enigh_nuevaconstruccion(
year = 2014, datos = "agro")
```

Figura 1  
Diagrama de relación de las bases de datos de la ENIGH (INEGI, 2014, p. 5)



### 3.4 Encuesta Nacional de Ocupación y Empleo

Es un proyecto estadístico de encuestas en hogares especializado en información sobre el mercado laboral. La ENOE provee datos trimestrales sobre

fuerza laboral, ocupación, subocupación y desocupación de los miembros del hogar encuestado. La Encuesta se compone de dos cuestionarios: el *socio-demográfico (CS)* y el *de ocupación y empleo (COE)*, este último en sus versiones básica y ampliada. La versión ampliada del COE se levanta en el trimestre I o II. Por lo tanto, cada levantamiento de la ENOE puede estar compuesto de hasta cinco bases de datos obedeciendo al tipo de instrumento de levantamiento del COE, que son las siguientes: 1) vivienda, 2) hogar, 3) CS, 4) COE parte I y 5) COE parte II.

La función `enoe()` descarga las bases de datos trimestrales de la ENOE desde el 2005. Su sintaxis tiene tres parámetros, de los cuales los dos primeros son indispensables. Primero, es necesario definir el año de levantamiento de los datos en formato numérico. Posteriormente, se debe especificar el trimestre del año en alfanumérico, que puede ser `tri m1`, `tri m2`, `tri m3` o `tri m4`. Por defecto, la función descarga por separado las cinco bases de datos que componen la Encuesta. Sin embargo, el tercero, `integrar`, que se describe al final de esta subsección, integra las cinco bases de datos en una sola matriz.

Cuando el parámetro `integrar` está (por *default*) en `FALSE`, la función descarga una lista de *R* con cinco bases de datos separadas. La función `summary` provee un resumen de estas (nombre, tamaño y tipo de objeto) y permite identificar la posición de cada una en la lista. En el siguiente bloque de código se extrae la base de datos del cuestionario de hogares de la ENOE 2005, trimestre I:

```
# Crear un objeto con las cinco bases de
# datos de la ENOE separadas
lista.enoe05 = enoe
(year = 2005, trimestre = "tri m1")
```

Esta lista, sin embargo, no es útil para el análisis estadístico; cada base de datos debe extraerse de la lista y almacenarse en un objeto de *R* distinto. Esto se logra, primero, identificando la posición de la base de datos en la lista con la función `summary()`, como se muestra en el siguiente bloque de código:

```
# Resumen de las cinco bases de
# datos
summary(lista.enoe05)
```

Posteriormente, se debe extraer la base de datos utilizando su posición dentro de la lista. En el siguiente bloque de código se ilustra la descarga de la base de datos *hogares*, que se encuentra en la tercera posición de la lista:

```
# Identificar y extraer base de
# datos del cuestionario
# hogares dt.hog105 = lista.enoe05[[3]]
```

Como se explicó, el parámetro por defecto `integrar = FALSE` descarga las bases de datos por separado, sin embargo, estas pueden integrarse directamente en una sola matriz utilizando la opción `integrar = TRUE`. Esta opción las conecta a todas mediante el identificador único que tienen en común. El siguiente bloque de código ilustra la descarga e integración de las bases de datos de la ENOE trimestre I:

```
# Descargar e integrar datos
# de la ENOE Trimestre I
lista.enoe.integrada = enoe (year =
2005, trimestre =
"tri m1", integrar = TRUE)
```

### 3.5 Marco Geoestadístico Nacional

Es un proyecto que presenta información sobre la división geoestadística<sup>2</sup> del país, así como otras formas de clasificación del territorio nacional.

La función `sig_marcogeo()` descarga los mapas del MGN desde 1995 hasta el 2013. Utiliza el paquete `rdgal` para descargar y gestionar archivos de información georreferenciada. Se puede acceder

<sup>2</sup> En sucesivos niveles de desagregación. Esta división está dada por los llamados *límites geoestadísticos*, que pueden coincidir con los político-administrativos oficiales, los cuales tienen sustento legal; sin embargo, los que no cuentan con este deben entenderse como provisionales (no tienen pretensión de oficialidad), trazados solo para realizar los operativos censales.



a la documentación en línea tecleando la función `sig_marcogeo()` (sin incluir nada entre los paréntesis). Su sintaxis tiene tres parámetros, de los cuales dos son indispensables. Primero, se debe especificar el año del mapa en formato numérico; los años disponibles se encuentran en la documentación de la función. El segundo es el mapa. El Marco provee cinco mapas diferentes: `entidades`, `municipios`, `ageb`, `urbano` y `rural`, aunque la disponibilidad de algunos varía entre años. El tercer parámetro, `version`, solo es necesario en los años que cuentan con múltiples versiones, por ejemplo, para el 2010 son 4.3, 5.0 o 5.0.A. Los años que requieren versión se especifican en la documentación de la función.

El mapa descargado debe ser almacenado en un objeto de *R*. El resultado es un objeto espacial de la clase `SpatialPolygonsDataFrame`, a partir del cual se puede realizar análisis espacial y estadístico. En el siguiente bloque de código se ejemplifica la descarga del mapa de estados del MGN de 1995 y se utiliza una función para su visualización:

```
# Descargar el MGN de 1995
mapa.estados95 <- sig_marcogeo
(year = 1995, mapa =
"entidades")
```

```
# Visualización del mapa
plot(mapa.estados95)
```

### 3.6 Red Nacional de Caminos

La RNC provee información georreferenciada sobre las vías de comunicación interurbana e intraurbana. Adicionalmente, contiene datos acerca de la infraestructura pública urbana (por ejemplo, túneles, puentes, plazas de cobro, marcas de kilometraje, etc.) y de la de otros medios de transporte (como transbordadores, aeropuertos, puertos y estaciones de ferrocarril).

Para utilizar los mapas de la RNC, se necesitan dos funciones. La primera, `sig_camionos_descarga()`, descarga los publicados entre el 2016 y 2019.

Esta función almacena en un objeto de *R* los 13 mapas que componen el conjunto de datos de la Red. En el siguiente bloque de código se ilustra la descarga del conjunto de mapas de la RNC para el 2019:

```
# Descargar mapas de la RNC de 2019
mapas.rnc <- sig_camionos_descarga
(year = 2019)
```

La ventaja de la función `sig_camionos_descarga()` es que descarga y almacena los 13 mapas de la RNC en una lista de objetos espaciales de *R*; sin embargo, la desventaja es que esta lista no es útil para el análisis espacial; cada mapa debe extraerse de la lista y almacenarse en un objeto espacial de *R* distinto. La segunda función, `sig_camionos_extrae()`, cumple este propósito.

Para extraer cada mapa, es necesario especificar, primero, el objeto con la lista de mapas previamente creado con `sig_camionos_descarga()`. En el segundo parámetro se debe especificar el mapa que se desea extraer de la lista. Los 13 mapas disponibles son: `estructura`, `localidad`, `manioobra_prohibida`, `plaza_cobro`, `poste_referencia`, `puente`, `red_vial`, `sitio_de_interes`, `tarifas`, `transbordador`, `red_localidad`, `red_sitio_de_interes` y `union`. También, la lista se puede consultar utilizando la función `summary()` sobre el objeto creado a partir de `sig_camionos_descarga()`. Enseguida se ilustra la extracción del mapa `puentes`, que contiene información de la ubicación geográfica de puentes carreteros:

```
# Obtener un mapa de puentes
mapa.puentes <- sig_camionos_extrae
(mapas.rnc, mapa = "puente")
```

## 4. Conclusiones y agenda de desarrollo

La provisión de herramientas para extender el acceso y procesamiento de datos para la investiga-

ción es una práctica que contribuye a mejorar la *reproducibilidad* de la ciencia y el acceso y uso de los datos abiertos (Peng, 2011; Munafò *et al.*, 2017). También, coadyuva a mejorar la transparencia, estandarización y automatización del proceso de investigación, desde la recolección de datos en bruto hasta la publicación final del producto.

El presente artículo busca apoyar tanto a la comunidad académica como a la profesional con una herramienta que facilite la obtención y el procesamiento de datos claves para la comprensión de la realidad nacional. El paquete *importinegi* provee funciones para acceder a datos abiertos del repositorio institucional del INEGI de manera simplificada, estandarizada y automatizada, con lo cual se quiere contribuir a mejorar el proceso en la investigación y enseñanza basadas en datos del Instituto. Fue desarrollado siguiendo las mejores prácticas de programación de paquetes en *R* (Wickham, 2015), *open source* y *reproducibilidad* de la ciencia (Gandrud, 2016).

La versión 1.0.0 de *importinegi*, expuesta en este documento, representa la primera fase de desarrollo del paquete, que busca seguir incorporando el acceso a más proyectos estadísticos del INEGI. La utilización de este *software* en la labor académica brinda la posibilidad de adoptar protocolos de *reproducibilidad* en el campo de las ciencias sociales, proporcionando a cada usuario el acceso a las bases de datos originales e íntegras producidas por el Instituto. Asimismo, al automatizar algunos procesos de descarga y procesamiento, reduce las capacidades de programación necesarias para utilizar las bases de datos del INEGI en el *software R*. Esto permite a las personas tanto del sector académico como profesional acceder y procesar bases de datos de amplia importancia para el análisis y la toma de decisiones.

## Fuentes

- Arel-Bundock, V. *Wdi: World development indicators (world bank)*. Vol. 2, 2013.
- Boettiger, C.; S. Chamberlain; E. Hart & K. Ram. "Building software, building community: lessons from the ropensci project", en: *Journal of open research software*. 3 (1), 2015.
- Gandrud, C. *Reproducible research with r and r studio*. Chapman and Hall/CRC, 2016.
- Gheri, F. & M. C. Kao. "Download and harmonize faostat and wdi data: the faostat package" en: *R Package Version*, 1. 2014.
- INEGI. *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2014*. México, INEGI, 2014.
- Lahti, L.; J. Huovari; M. Kainu & M. Biecek. "Retrieval and analysis of eurostat open data with the eurostat package", en: *The R Journal*. 9 (1), 2017, pp. 385-392.
- Munafò, M. R.; B. A. Nosek; D. V. Bishop; K. S. Button; C. D. Chambers; N. P. du Sert & J. P. Ioannidis. "A manifesto for reproducible science" en: *Nature Human Behaviour*. 1 (1), 0021, 2017.
- Peng, R. D. "Reproducible research in computational science", en: *Science*. 334 (6060), 2011, pp. 1226-1227.
- Walker, K. *tidycensus: Load us census boundary and attribute data as 'tidyverse' and 'sf'-ready data frames. r package version 0.4*. 1. 2018.
- Wickham, H. *R packages: organize, test, document, and share your code*. O'Reilly Media, Inc., 2015.
- Labo. Alibus, sam et odia quam, nobis porest, te nis imus di alignim pellect atenducium quo vendio omnim adictibusam iusandi oditium harioeribus eos earciis sae cuptatae eos dolore peritiones re invol ipsandae. Odis ertestio. Atiis eturistius dolendis santo quibusto excaequa perit maio experupitat.
- Lor rehendis comnis im exped quodist, que vel ex eos dolupta ssitibustium nonsera escuscienet volor alias aruptaq uiscis suntis si ut ut laborum dem ipsam quunt.
- Occus qui consent orporporem rentem que vendi aut landit ex estrum, verio es nos ea pos molenim olorist, to magniscimi, tendige ndusapitem evellenimet renda simendam quae nos entorrum fuga. Itatem nati aspe laut estotam fugit officipsam as errum que pa ipit fuga. Henimposant quia des sapit quam vel is et quam, consequamus eat autem ilit, officae. Tiatem facerunt im ea venihil iur?