

El programa R: una estrategia inicial para su entendimiento y aprendizaje

R program: an initial strategy for understanding and learning it

Francisco Joel Jahuey Martínez, Jessica Beatriz Herrera Ojeda y
Francisco Alejandro Paredes Sánchez

Resumen

R es un lenguaje de programación estadístico libre y es la plataforma más utilizada en el área del análisis de datos. Debido a las diversas bondades que ofrece este programa, es utilizado en muchas ciencias y es la fuente de análisis de otros softwares. El programa R puede ser explorado mediante diversas plataformas como: R, RStudio, RStudio Cloud y Microsoft R Open. Su instalación es rápida, sencilla y gratuita. Aprender a utilizar R requiere del entendimiento de conceptos básicos como son las funciones y objetos, por lo que conocer su estructura, uso y desarrollo ayudarán al lector a explorar todas las bondades y ventajas que este software ofrece. En este artículo se explican los conceptos básicos en el entorno de R y se proporcionan diferentes opciones para que cualquier usuario, con o sin experiencia en lenguajes de programación, logre utilizar este maravilloso programa estadístico.

Palabras clave: R, software estadístico, funciones básicas, objetos, bases de datos.

Abstract

R is a free statistical programming language and is the most widely used platform in the area of data analysis. Due to its various benefits, this program it is used in many sciences and is the source of analysis of other software. The R program can be explored through various platforms such as: R, RStudio, RStudio Cloud and Microsoft R Open. Installation is quick, easy and free. Learning to use R requires the understanding of basic concepts such as functions and objects, so knowing its structure, use and development will help the reader to explore all the benefits and advantages that this software offers. This article explains the basic concepts over the R environment and provides different options, so that any user, with or without experience in programming languages, can use this wonderful statistical program.

Keywords: R, statistical software, basic functions, objects, databases.

CÓMO CITAR ESTE TEXTO

Jahuey Martínez, Francisco Joel, Herrera Ojeda, Jessica Beatriz y Paredes Sánchez, Francisco Alejandro. (2022, julio-agosto). El programa R: una estrategia inicial para su entendimiento y aprendizaje. *Revista Digital Universitaria (RDU)*, 23(4). <http://doi.org/10.22201/cuaieed.16076079e.2022.23.4.4>

Francisco Joel Jahuey Martinez

Facultad de Zootecnia y Ecología, Universidad Autónoma de Chihuahua (FZYE-UACH)

Licenciado en Biología por el Instituto Tecnológico de Ciudad Victoria (ITCV). Cursó la Maestría y Doctorado en el Centro de Biotecnología Genómica del Instituto Politécnico Nacional (CBG-IPN). Realizó estancias académicas en la Iowa State University (ISU) de Estados Unidos y en la University of New England (UNE) de Australia. Actualmente es miembro del Sistema Nacional de Investigadores y profesor con perfil PRODEP en la FZYE-UACH. Su principal línea de investigación se enfoca en el análisis bioinformático de datos biológicos, predicción genómica y mapeo genético mediante GWAS en especies de importancia doméstica.

 fjahuey@uach.mx

 orcid.org/0000-0002-6562-5875

 Francisco-Jahuey

Jessica Beatriz Herrera Ojeda

Instituto Tecnológico del Valle de Morelia (ITVM)

Médico Veterinario Zootecnista por la Universidad Autónoma de Yucatán (UADY) y Doctora en Ciencias Biológicas por la Universidad Michoacana de San Nicolás de Hidalgo. Realizó una estancia posdoctoral en el Centro de Biotecnología Genómica del Instituto Politécnico Nacional. Actualmente es miembro del Sistema Nacional de Investigadores y profesora investigadora en el ITVM.

 jessica.ho@vmorelia.tecnm.mx

 orcid.org/0000-0001-5293-1550

 Jb_Herrera-Ojeda

Francisco Alejandro Paredes Sánchez

Unidad Académica Multidisciplinaria Mante, Universidad Autónoma de Tamaulipas (UAMM-UAT)

Ingeniero en Biotecnología por la Universidad Politécnica de Zacatecas y Maestro en Ciencias en Biotecnología Genómica y Doctor en ciencias en Biotecnología por el Centro de Biotecnología Genómica del Instituto Politécnico Nacional. Ha desarrollado líneas de investigación sobre el análisis genético molecular de bovinos para la identificación de marcadores moleculares asociados a características productivas e identificación de marcadores moleculares asociados a resistencia a insecticidas en plagas de interés agrícola. Perfil PRODEP 2020-2023. Docente-Investigador en la UAMM-UAT.

 faparedes@docentes.uat.edu.mx

 orcid.org/0000-0003-2661-6607

 Francisco-Paredes-Sanchez

¿Qué es R?

R es un lenguaje de programación utilizado como software estadístico en muchas disciplinas. Al ser gratuito y fácil de utilizar se ha convertido en la plataforma estadística más utilizada para el análisis de datos y big data; está a la altura de otros lenguajes y programas estadísticos como [Python](#), [Julia](#), [SPSS](#) y [SAS](#).

Al ocupar la estadística un lugar importante en la investigación científica, aprender R representa una excelente competencia académica y laboral en el entorno actual del *big data*. Por lo tanto, este artículo pretende instruir al lector en el entendimiento y manejo básico del programa R, a través de una serie de pasos y recomendaciones que lo guiarán en el proceso de aprendizaje, a partir de una mínima o nula experiencia en programación.

¿Cómo descargo el programa?

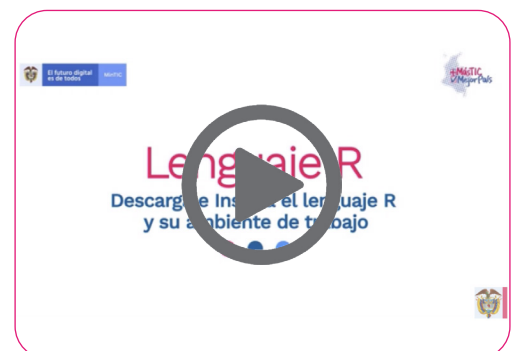
Para empezar, es importante saber que R cuenta con varias plataformas de uso como: [RStudio](#), [RStudio Cloud](#) y [Microsoft R Open](#), además de una serie de herramientas accesorias y

repositorios que se explicarán más adelante (ver figura 1). La consola natural puede ser descargada desde el [sitio web oficial de R](#). El proceso de instalación es muy fácil (ver video 1), basta únicamente con seleccionar la versión adecuada para el sistema operativo de nuestra computadora y seguir los pasos de instalación que vienen definidos por defecto. La plataforma de R está disponible para los sistemas operativos Windows, MacOS y Linux, y no requiere de grandes cantidades de memoria RAM para su uso básico, por lo que puede ser utilizada incluso en minilaptops.

Por otro lado, RStudio es una plataforma más llamativa y amigable para el usuario ya que contiene múltiples menús y opciones para el manejo de datos. RStudio Cloud por otra parte resulta interesante para los usuarios que no desean tener instalado R en su ordenador. Para utilizar RStudio Cloud solo se necesita crear una cuenta [aquí](#) y luego el sistema proporciona una versión en línea de RStudio bajo ciertas restricciones de uso (1 Gb de RAM, 1 CPU y 15 proyectos como límite). Como cuarta opción se encuentra Microsoft R Open que es una plataforma de R desarrollada por Microsoft que incluye capacidades adicionales para mejorar el rendimiento y la reproducibilidad del programa.

Figura 1. Diferentes plataformas para el uso del programa R y principales repositorios de librerías de R.

Video 1. Cómo descargar e instalar R (Ministerio TIC Colombia, 2020).



Cuadro 1. Algunas funciones básicas que puedes emplear con R

Función	Grupo	¿Qué hace?
<i>getwd</i>	Exploración de entorno	Obtiene la carpeta o directorio de trabajo actual
<i>setwd</i>	Exploración de entorno	Cambia el directorio de trabajo
<i>list.files</i>	Exploración de archivos	Conocer el contenido de tu directorio de trabajo
<i>read.table</i>	Exploración de archivos	Lee cualquier tipo de archivo que contenga una tabla
<i>rbind</i> y <i>cbind</i>	Exploración de datos	Unir tablas por filas o por columnas
<i>head</i> y <i>tail</i>	Exploración de datos	Visualizar partes de un archivo, tabla, matriz
<i>sum</i>	Análisis de datos	Sumar
<i>t</i> , <i>solve</i> , <i>det</i>	Análisis de datos	Realizar cálculos con matrices

Buscar funciones de R

Para saber cuál función debe emplear el usuario, primero requiere definir la tarea que desea ejecutar; por ejemplo, sumar por filas, filtrar con base en varias columnas, resumir múltiples variables, generar gráficas, etcétera. Un primer intento para encontrar la función adecuada es realizar una búsqueda en Google utilizando la famosa frase *¿how to ... in R?*. Por ejemplo: *¿how to read a file in R?* y *¿how to filter a dataframe in R?*, que se refieren a *¿cómo leer un archivo en R?* y *¿cómo filtrar una tabla en R?*, respectivamente. Esta búsqueda arrojará respuestas en sitios como [rdocumentation](#) y [statmethods](#). Bastará con leer la descripción de la función proporcionada en el sitio web y si esta función cumple las necesidades del usuario, simplemente se debe seguir el ejemplo de ejecución, cumpliendo siempre con la estructura de datos solicitada por la función.

Otro as bajo la manga que tiene R para identificar una función es definir la orden que se desea ejecutar en forma de verbo y traducirlo al inglés, ya que los nombres de algunas funciones de R son simples expresiones comunes como **paste** y **summary**, que realizan las tareas

de pegar y resumir, respectivamente. En otros casos el usuario notará que el nombre de la función es una abreviación de la tarea que realiza; por ejemplo, para obtener o definir el directorio de trabajo: **getwd** (get working directory) y **setwd** (set working directory).

Cuando ya nos encontremos más cómodos con R, podríamos tratar de realizar tareas más complejas, que involucren múltiples procedimientos y el uso de más de una función. Para ello se recomienda aplicar la misma estrategia de búsqueda y apoyarse en sitios web como [stackoverflow](#) y [biostars](#), que son foros donde se discuten preguntas entorno al uso del programa R. Aquí se podrán encontrar diferentes soluciones al problema en cuestión. Sólo se deberán copiar los scripts¹ y posiblemente analizar detenidamente cada paso del código.

Ya encontré la función, y ahora ¿cómo la utilizo?

Recordemos que utilizar R implica “hablar idioma computadora”, por lo que un paso crucial al aprender R es saber que las funciones se deben escribir bajo un orden específico o *sintaxis*, para que puedan trabajar correctamente. La mayoría de las

¹ Secuencia de comandos para resolver una o más tareas

funciones se ejecutan escribiendo el nombre de la función seguido de paréntesis que contiene el nombre del objeto al que se desea aplicar la función. Por ejemplo, para que podamos sumar los datos ponemos la siguiente función: *sum(datos)*. En caso de no utilizar una sintaxis correcta o de no escribir bien el nombre de la función o del objeto, surgirá el mensaje más molesto para todo usuario de R, el famoso “Error”.

Además, en ocasiones, será necesario darle algunas instrucciones extra a R para que la función haga exactamente lo que nosotros queremos que haga. Para esto se deben definir una serie de argumentos, que determinan el comportamiento de las funciones. Por ejemplo, en la función **order** (ordenar), debemos aumentarle el argumento **decreasing** (para definir en qué sentido), si queremos que nos ordene los datos de manera descendente.

Algunas funciones pueden tener uno o más argumentos. Afortunadamente, muchos de ellos están definidos por defecto, por lo que el usuario solo deberá indicarlos o modificarlos si se quiere ajustar algún parámetro. Para saber qué argumentos están definidos en una función se puede aplicar el comando **args** a la función de interés, o bien, revisar la descripción de la función, para esto R cuenta con el comando **“?”**, que al escribirlo seguido del nombre de la función despliega una ventana con los datos descriptivos de la función intente *args(mean)* o *?mean*.

En R, la escritura y ejecución de códigos se puede hacer de manera

continua, es decir, el resultado de una función se puede utilizar como dato de entrada de otra función. Por ejemplo, cuando se aplica la función **table** para describir una variable cualitativa, el resultado o salida es un conteo de frecuencias, que puede pasarse directamente a la función **pie**, que generará un gráfico de pastel sin la necesidad de crear un objeto que contenga las frecuencias.

A medida que el usuario adquiera experiencia y dominio sobre las funciones de R podrá escribir instrucciones cada vez más complejas. Por supuesto, al principio tenderá a escribir códigos demasiado extensos, aunque funcionales y entendibles, pero conforme aumente su dominio sobre el lenguaje, comenzará a utilizar funciones y procedimientos más eficientes y cortos. Como recomendación, antes de empezar a escribir códigos en R, vale la pena trazar un mapa detallado de todos los procedimientos que se desean realizar sobre la base de datos, e ir resolviendo paso a paso cada una de las tareas. De este modo el usuario podrá notar un avance en sus análisis y evitará caer en la confusión o frustración. Con el paso del tiempo, el usuario notará que sus procedimientos serán cada vez más metódicos y ordenados. Aquí, el uso de editores de texto o de herramientas como [Tinn-R](#) sirven de apoyo en la escritura de códigos.

¿Cómo genero los objetos?

En R, los objetos son las estructuras que sirven para almacenar datos y deben ser creados por el usuario. Existen cinco tipos de estructuras básicas que son: vectores, tablas o

dataframes, matrices, listas y arreglos, cada uno con atributos específicos. Estos objetos pueden ser creados mediante las funciones **vector**, **data.frame**, **matrix**, **list** y **array**.

El tipo de objeto más utilizado es el **dataframe**, que es una estructura de datos en forma de tabla, donde las filas representan a las observaciones y las columnas a las variables. De hecho, la mayoría de las funciones básicas de R que sirven para leer y cargar archivos, como **read.table** o **read.csv**, generan un objeto tipo dataframe. También su creación es muy simple, basta con indicar el nombre y el contenido de cada columna, por ejemplo, para guardar información dentro del objeto denominado "grupos" se utiliza: `grupos <- data.frame (individuo= c(1:5), grupo= c("A","B","B","C","A"))` y como resultado se obtiene una tabla de 5 filas (1 por cada individuo) y 2 columnas (la de individuo y la de grupos). Note que para guardar información dentro de un objeto se debe utilizar el operador de asignación "**<-**", que es una especie de flecha que indica en dónde se depositarán los datos de interés (ver figura 4).

Independientemente de la manera en que se generen los objetos,

todos deben tener un nombre que los identifique. Aunque se pueden nombrar de muchas maneras, se recomienda que el nombre de los objetos sea corto y alusivo a su contenido, para que el usuario pueda identificarlos fácilmente. Esto es muy importante ya que durante del proceso de aprendizaje, se estarán creando muchos objetos. Aquí la función **ls** es de gran utilidad ya que ayuda a conocer todos los objetos presentes en una sesión de R.

¿Cómo puedo practicar R si no tengo una base de datos?

Por lo general, el primer procedimiento en el análisis de datos será leer o cargar archivos propios a una sesión de R. Sin embargo, muchas veces los aprendices de R no cuentan con datos o archivos para practicar. Por fortuna, R cuenta con la librería "datasets", que contiene datos de diferente índole para su uso y exploración. Además, la mayoría de las librerías especializadas de R incluyen bases de datos, por lo que si el usuario ha elegido alguna deberá seguir, en primer lugar, el tutorial de uso de la librería y, luego, estructurar sus datos siguiendo el

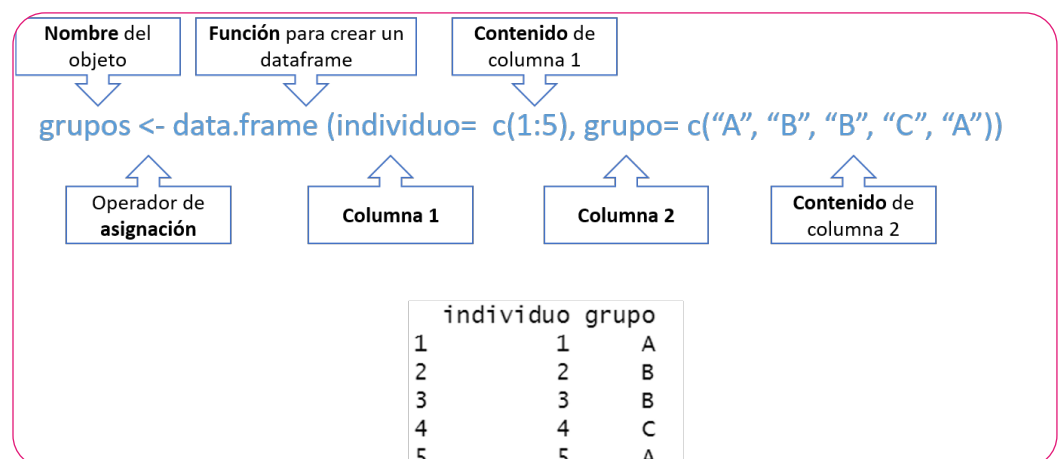


Figura 4. Creación de un objeto tipo dataframe.

mismo formato del ejemplo. R también cuenta con funciones para simular variables aleatorias con diferente tipo de distribución, por ejemplo, uniforme (**runif**), normal (**rnorm**) y binomial (**rbinom**), aunque recomendamos que éstas sean usadas por usuarios con conocimientos más avanzados tanto en R como en estadística.

¿Cómo puedo agilizar y automatizar procedimientos? Loops en R

En ocasiones el usuario necesitará repetir muchas veces un mismo procedimiento y posiblemente incluir condicionales dentro de ellos, es decir, buscará automatizar sus análisis. En R, esto se puede hacer mediante procedimientos denominados *loops* y a través de las funciones **for**, **while** y **repeat**². Aunque parecen simples, estas funciones son de gran utilidad en el análisis de datos, ya que permiten realizar múltiples procedimientos en una sola instrucción.

Por ejemplo, para crear loops mediante la función **for**, el usuario debe indicar el iterador (generalmente la letra “i”) seguido de un vector sobre el cual se realizarán las iteraciones. Luego, entre llaves “{}” deberá indicar el conjunto de procedimientos que desea ejecutar y que representarán el cuerpo del loop, por ejemplo: *for(i in 1:5) {print (i^2)}*. En el ejemplo, cada valor del 1 al 5 será elevado al cuadrado. La ventaja de hacerlo mediante un loop es que no tenemos que escribir la operación 5 veces, sino que se le da la indicación al programa en una sola instrucción.

A pesar de su gran utilidad, los loops también representan uno de los principales dolores de cabeza para todos los principiantes en R, ya que su entendimiento y utilización requieren de mucha práctica. Algunas recomendaciones para aprender a ejecutar loops son: iniciar practicando con la función **for**; utilizar pocas instrucciones dentro del cuerpo del loop, e imprimir o guardar el resultado de cada iteración. También se recomienda que antes de todo se ejecuten por separado las instrucciones del cuerpo del loop, para asegurarse de que trabajan correctamente y que realicen la tarea deseada.

Aunque los principiantes en R tratarán de evitar los loops, la mayoría terminarán aprendiéndolos de manera forzada, ya que en algún momento del análisis de datos se enfrentarán a un problema real donde deberán aplicarlos. Por experiencia, los usuarios que logran escribir y ejecutar loops pronto se vuelven independientes en el uso del programa R, son capaces de seguir tutoriales sin ningún problema, con lo que comienzan el proceso del autoaprendizaje, y desarrollan sus propias funciones en R.

Otras maneras de aprender R

Otra manera de aprender R es mediante la librería **Rcmdr**, que a través de una interfaz gráfica facilita la ejecución de operaciones en R sin la necesidad de introducir código. La ventaja de utilizar esta **Rcmdr** es que el usuario puede observar los comandos ejecutados por la librería cada vez que se realiza algún procedimiento

²El comando **for** se utiliza para repetir cualquier instrucción tantas veces como el usuario lo desee. El comando **while** realiza la misma operación, pero la instrucción se ejecutará siempre y cuando se cumpla una condición predefinida. De manera similar, el comando **repeat** también repetirá los procedimientos indicados por el usuario, pero en este caso la instrucción se detendrá hasta que se cumpla cierta condición.

gráfico o estadístico, desde la ventana de interfaz gráfica. Otra opción es mediante la librería **swirl**, que a través de ejercicios y preguntas de opción múltiple enseña de manera interactiva los conceptos básicos para el análisis de datos. La librería **swirl** contiene cursos de aprendizaje de R para nivel principiante, intermedio y avanzado. También existen los denominados *bookdowns*, que son libros electrónicos, muy explicativos, que incluyen códigos para practicar R. La figura 5 muestra algunos disponibles en la web y una lista completa se encuentra disponible [aquí](#). Por supuesto, videos tutoriales también pueden ser de gran ayuda en el aprendizaje de R. La última opción, pero no menos importante, es el seguimiento de cursos ofertados por universidades o mediante plataformas educativas como [datacamp](#), [udemy](#), [coursera](#), etcétera.

surgirán errores de código que provocarán frustración en la mayoría de los usuarios, pero entre más se practique este nuevo lenguaje mejor será el desempeño, confianza y agilidad en el uso del programa, y como resultado se desarrollarán nuevas habilidades para el análisis de datos. Una vez que se logre tener un dominio básico del programa, el usuario de R será capaz de instalar el programa y sus diferentes librerías, de identificar y utilizar funciones básicas, y de construir secuencias lógicas de procedimientos para su ejecución.

Te invitamos a seguir utilizando este software y que logres llegar al siguiente paso: ¡el nivel intermedio! Te recomendamos practicar los conceptos básicos de estadística, e iniciar con el uso de funciones descriptivas, funciones gráficas, hasta pasar por los conceptos de estadística inferencial.

Reflexiones

Aprender el lenguaje R será como aprender un nuevo idioma. Al principio,

Figura 5. *Bookdowns* para el aprendizaje de R disponibles en la web.



Referencias

- ❖ Ministerio TIC Colombia. (2020, febrero 3). *Instalación Lenguaje R* [Video]. YouTube. https://youtu.be/bl_rwoDw4f4

Sitios de interés

- ❖ [Bookdown: R para principiantes de Juan Bosco Mendoza Vega](#)
- ❖ [Funciones básicas de R \(Video\)](#)
- ❖ [¿Qué puede hacer el software R para resolver tus problemas?, Rodríguez Silva, Revista Digital Universitaria](#)
- ❖ [El Proyecto R para estadística computacional](#)

Recepción: 09/08/2021. Aprobación: 18/05/2022.