

Introducción al análisis de datos con R

Daniel Blanc

10/31/2022

Table of contents

Prefacio

Sobre este ebook

El objetivo de este ebook es brindarle al lector una introducción amigable pero completa sobre el manejo básico de R para el análisis de datos. Se presupone que quien utilice este manual desde el comienzo, no tenga conocimientos previos de R o de programación, por tanto, podrá utilizar esta herramienta para dar sus primeros pasos en el mundo de la ciencia de datos.

También puede ser de utilidad para quien tenga experiencia, en la medida que cada tanto es de utilidad repasar los conceptos fundamentales, especialmente si hace tiempo que no realizamos determinada tarea, o si nunca se tuvo la necesidad de encarar determinado proyecto, por ejemplo, una Shiny App.

El libro se divide en X capítulos:

1. Introducción a R, funciones y RStudio
2. Importación de datos
3. Estructura de dataframes
4. Análisis de datos con R
5. Manipulación de datos
6. Visualización
7. Informes interactivos y automatizados
8. Dashboards interactivos

Sobre mi

Soy un analista de datos con experiencia en proyectos de analytics y machine learning aplicado a negocios. Actualmente me desempeño como analista de datos senior en el [Instituto Nacional de Estadística](#) y he colaborado en el desarrollo de modelos predictivos para la consultora BCS y en la elaboración de dashboards como consultor independiente. He sido docente de Python, R y Power BI en [Instituto CPE](#) durante cuatro años, donde también me desempeño



como coordinador de proyectos. Finalmente, también fui docente de desarrollo web para [Plan Ceibal](#).

1 Introducción

Daniel Blanc

1.1 Interacción con R

Como bien lo describe su página web, **R** es un lenguaje desarrollado para análisis matemático-estadístico. En 1995 evoluciona de su antecesor, el lenguaje de programación S, y comienza su expansión por el mundo académico, la cual hoy complementa con la ciencia de datos. Junto con Python, es el más usado dentro del rubro.¹

Un lenguaje de programación interactúa con el usuario, generalmente, a través de una consola, en donde el usuario ingresa ciertas “órdenes” al lenguaje, y éste devuelve un resultado. El ejemplo más sencillo es si usáramos R como una calculadora.

```
4 + 4
```

```
[1] 8
```

Como puede apreciarse, yo envíé una orden a R: que sume $2 + 2$, y éste me devuelve el resultado. Además de esto, podría haber almacenado dicho resultado dentro de un objeto. Los objetos en R son básicamente elementos que guardan información que nosotros les depositamos dentro, y que tenemos que darles un nombre para luego referirnos a ellos. Por ejemplo, guardaré el resultado de esta suma dentro de un objeto que se llame “resultado_suma”.

```
resultado_suma <- 2 + 2
```

Varias cosas: para empezar, puede apreciarse que para asignar un resultado a un objeto, debemos utilizar el “<-”, o *asignador*. Esto refiere a que básicamente todo lo que esté a la derecha del símbolo será almacenado en el objeto de la izquierda. Por otro lado, ¿por qué usé un guión bajo para nombrar a mi objeto? Es que **R no permite espacios en nombres**

¹Para ver más sobre la relación de R y Python ir a [este artículo](#).

```
Rterm (64-bit)

R version 4.1.3 (2022-03-10) -- "One Push-Up"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> _
```

de objetos, tampoco tildes ni caracteres especiales fuera de los básicos como el guión bajo. Se recomienda sustituir los espacios por guiones o mayúsculas. Finalmente vemos que no ha devuelto nada la consola, eso es porque se ha guardado el valor correctamente, **al almacenar valores la consola no devuelve ningún resultado, para acceder al mismo hay que llamar al objeto creado.**

```
resultado_suma
```

```
[1] 4
```

Finalmente, los objetos en R también interactúan entre ellos. Podemos generar distintos objetos y obtener resultados al generar su interacción. Por ejemplo, a continuación tomaré el objeto ya existente *resultado_suma* y le aplicaré una serie de interacciones.

Primero, veremos cómo R permite seguir interactuando con los objetos luego de haberlos creado.

```
resultado_suma + 10
```