Aprender R: iniciación y perfeccionamiento

François Rebaudo

2018-06-28

Contents

Chapter 1

Agradecimientos

Agradezco a todos los colaboradores que ayudaron a mejorar este libro con sus consejos, sugerencias de cambios y correcciones:

=> lista para actualizar en la publicación.

Las versiones de gitbook, html y epub de este libro usan los iconos de fuente abierta de Font Awesome (`https://fontawesome.com`). La versión en PDF utiliza los iconos del proyecto Tango disponibles en openclipart (`https://openclipart.org/`). Este libro fue escrito con el paquete R bookdown (`https://bookdown.org/`). El código fuente está disponible en GitHub (`https://github.com/frareb/myRBook_SP`). La compilación usa Travis CI (`https://travis-ci.org`). La versión en línea se aloja y actualiza a través de Netlify (`http://myrbooksp.netlify.com/`).

Chapter 2

Licencia

Licencia Reconocimiento-NoComercial-SinObraDerivada 3.0 España (CC BY-NC-ND 3.0 ES ; `https://creativecommons.org/licenses/by-nc-nd/3.0/es/`)

Esto es un resumen inteligible para humanos (y no un sustituto) de la licencia.

Usted es libre de:

- Compartir — copiar y redistribuir el material en cualquier medio o formato.
- El licenciador no puede revocar estas libertades mientras cumpla con los términos de la licencia.

Bajo las condiciones siguientes:

- Reconocimiento — Debe reconocer adecuadamente la autoría, proporcionar un enlace a la licencia e indicar si se han realizado cambios<. Puede hacerlo de cualquier manera razonable, pero no de una manera que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.
- NoComercial — No puede utilizar el material para una finalidad comercial.
- SinObraDerivada — Si remezcla, transforma o crea a partir del material, no puede difundir el material modificado.
- No hay restricciones adicionales — No puede aplicar términos legales o medidas tecnológicas que legalmente restrinjan realizar aquello que la licencia permite.

Avisos:

No tiene que cumplir con la licencia para aquellos elementos del material en el dominio público o cuando su utilización esté permitida por la aplicación de una excepción o un límite. No se dan garantías. La licencia puede no ofrecer todos los permisos necesarios para la utilización prevista. Por ejemplo, otros derechos como los de publicidad, privacidad, o los derechos morales pueden limitar el uso del material.

Chapter 3

introducción

3.1 Por qué apprender R

3.2 Este libro

3.3 Lectura adicional en español

- R para Principiantes, Emmanuel Paradis (`https://cran.r-project.org/doc/contrib/rdebuts_es.pdf`)
- xxx

Part I

Conceptos básicos

Chapter 4

Primeros pasos

4.1 Instalar R

El programa para instalar el software R se puede descargar desde el sitio web de R: `https://www.r-project.org/`. En el sitio web de R, primero es necesario elegir un espejo CRAN (servidor desde el que descargar R, el más cercano a su ubicación geográfica), luego descargue el archivo *base*. Los usuarios de Linux pueden preferir un `sudo apt-get install r-base`.

El software R se puede descargar de muchos servidores CRAN (Comprehensive R Archive Network) de todo el mundo. Estos servidores se llaman espejos. La elección del espejo es manual. Información adicional como esta nota siempre estará representada con este pictograma *información*.

4.2 R como calculadora

Una vez que se inicia el programa, aparece una ventana cuya apariencia puede variar dependiendo de su sistema operativo (Figura 4.1). Esta ventana se llama *consola*.

La consola corresponde a la interfaz donde se interpretará el código, es decir, donde el código será transformado en lenguaje de máquina, ejecutado por la computadora y retransmitido en forma legible por humanos. Esto corresponde a la pantalla de una calculadora (Figura 4.2). Así es como se usará R más adelante en esta sección.

A lo largo de este libro, los ejemplos del código R aparecerán sobre un fondo gris. Se pueden copiar y pegar directamente en la consola, aunque es mejor reproducir escribiendo los ejemplos en la consola (o más adelante en los scripts). El resultado de lo que se envía en la consola también aparecerá en un fondo gris con ## delante del código para hacer la distinción entre el código y el resultado del código.

```
5 + 5
```

```
## [1] 10
```

Si escribimos `5 + 5` en la consola y luego `Enter`, el resultado aparece precedido por el número [1] entre corchetes. Este número corresponde al número del resultado (en nuestro caso, solo hay un resultado, volveremos a este aspecto más adelante). También podemos observar en este ejemplo el uso de espacios antes y después del signo +. Estos espacios no son necesarios,

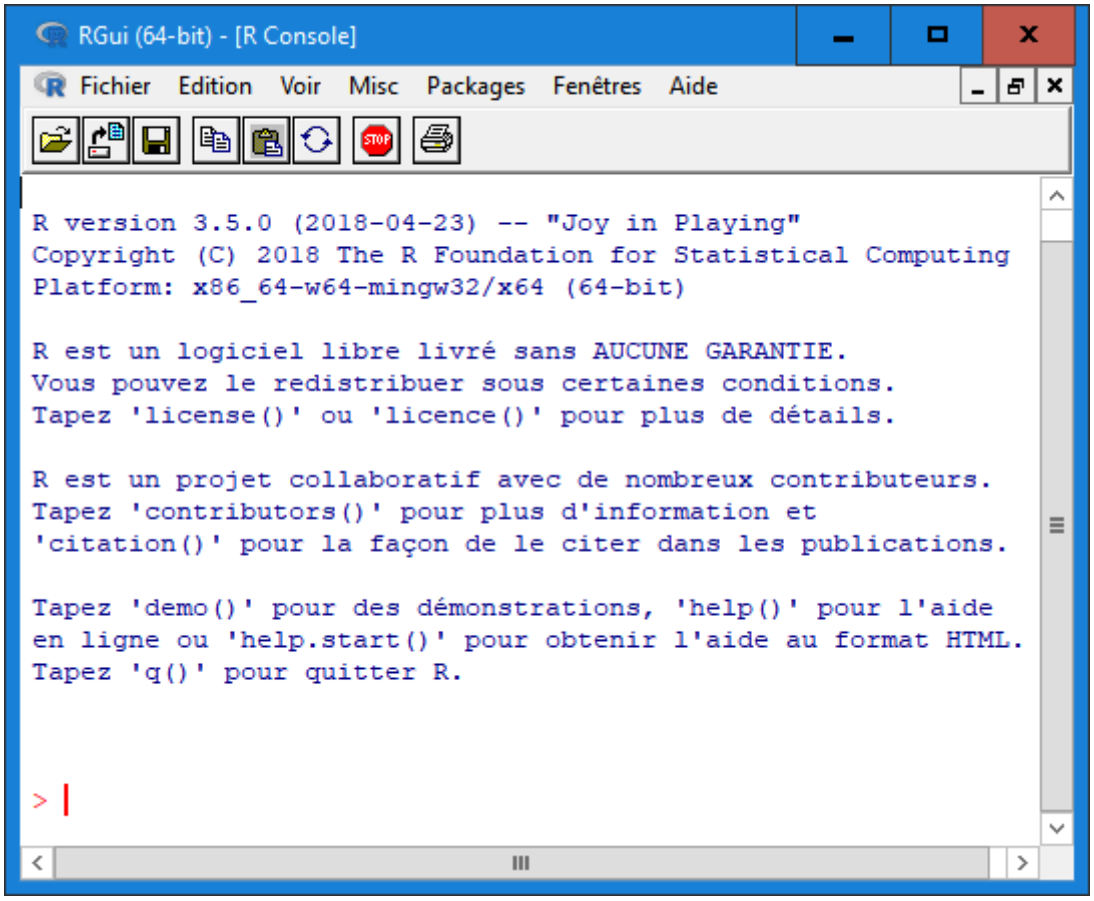

Figure 4.1: Captura de pantalla de la consola R en Windows.

Table 4.1: Operadores aritméticos.

Label	Operador
adición	+
resta	-
multiplicación	*
división	/
potencia	^
módulo	%%
cociente decimal	%/%

pero permiten que el código sea más legible para los humanos (es decir, más agradable de leer tanto para nosotros como para las personas con las que queremos compartir nuestro código). Los operadores aritméticos disponibles en R se resumen en la tabla 4.1.

Clásicamente, las multiplicaciones y divisiones tienen prioridad sobre las adiciones y sustracciones. Si es necesario, podemos usar paréntesis.

```
5 + 5 * 2
```

```
## [1] 15
```

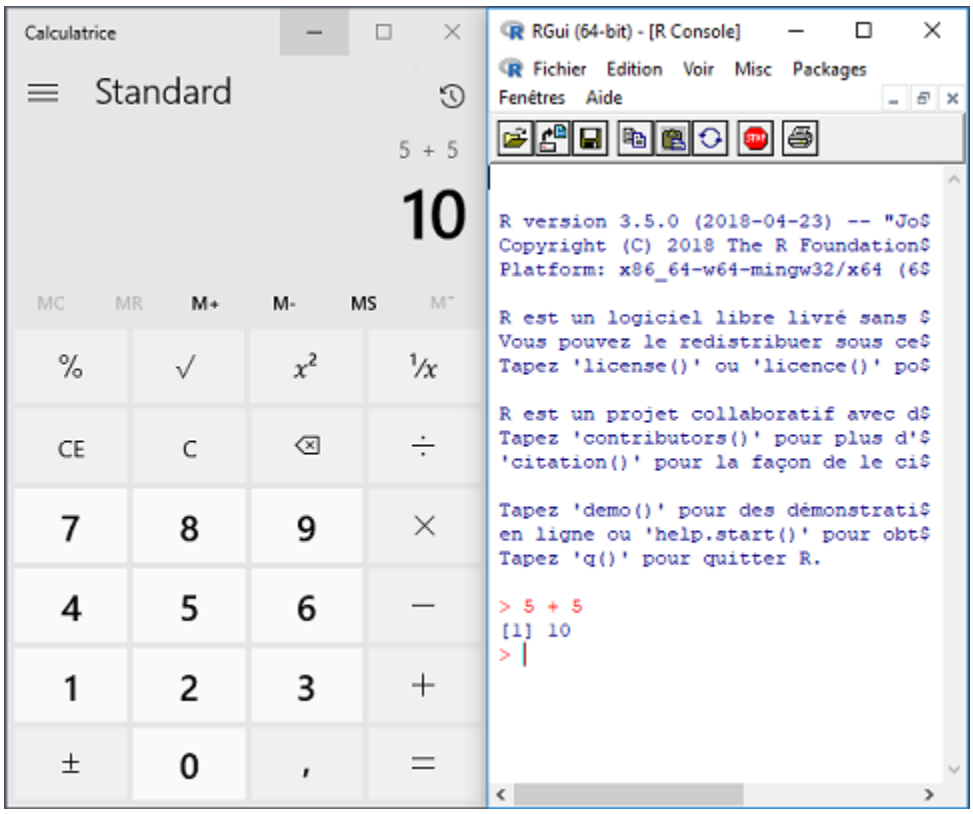

Figure 4.2: Captura de pantalla de la consola R al lado de la calculadora de Windows.

```
(5 + 5) * 2
```

```
## [1] 20
```

```
(5 + 5) * (2 + 2)
```

```
## [1] 40
```

```
(5 + 5) * ((2 + 2) / 3)^2
```

```
## [1] 17.77778
```

El operador de módulo corresponde al resto de la división euclidiana. Se usa en ciencias de la computación, por ejemplo, para saber si un número es par o impar (un número módulo 2 devolverá 1 si es impar y 0 si es par).

```
451 %% 2
```

```
## [1] 1
```

```
288 %% 2
```

```
## [1] 0
```

```
(5 + 5 * 2) %% 2
```

```
## [1] 1
```

```
((5 + 5) * 2) %% 2
```

```
## [1] 0
```

R también incorpora algunas constantes que incluyen `pi`. Además, el signo infinito está representado por `Inf`.

```
pi
```

```
## [1] 3.141593
```

```
pi * 5^2
```

```
## [1] 78.53982
```

```
1/0
```

```
## [1] Inf
```

el *estilo* del código es importante porque el código está destinado a ser leído por nosotros y por otras personas. Para tener un estilo legible, se recomienda colocar espacios antes y después de los operadores aritméticos, excepto "*", "/" y "^", aunque a veces es útil agregarlos como es el caso en nuestro ejemplos. La información sobre el *estilo* siempre estará representada con este pictograma para que sea fácilmente identificable.

4.3 El concepto de objeto

Un aspecto importante de la programación con R, pero también la programación en general es la noción de objeto. Como se indica en la página web de wikipedia (`https://ia.wikipedia.org/wiki/Objecto_(informatica)`), en ciencias de la computación, un objeto es un *contenedor*, es decir, algo que contendrá información. La inforamción contenida en un objeto puede ser muy diversa, pero por el momento contendremos en un objeto el número 5. Para hacer esto (y para reutilizarlo más adelante), debemos darle un nombre a nuestro objeto. Con R, los nombres de los objetos no deben contener caracteres especiales como *^ $? | + () [] {}*, entre otros. No deben comenzar con un número ni contener espacios. El nombre del objeto debe ser representativo de lo que contiene, sin ser demasiado corto ni demasiado largo. Imagine que nuestro número 5 corresponde al número de repeticiones de un experimento. Nos gustaría darle un nombre que se refiera a *numero* y *repeticiones*, que podríamos reducir a *nbr* y *rep*, respectivamente (*nbr* para number en ingles). Hay varias posibilidades que son bastante comunes bajo R:

- la separación mediante *guión bajo* (underscore): `nbr_rep`
- la separación mediante el carácter *punto*: `nbr.rep`
- el uso de letras minúsculas: `nbrrep`
- el estilo *lowerCamelCase* que consiste en una primera palabra en minúscula y la primera letra de las siguientes palabras con una letra mayúscula: `nbrRep`
- el estilo *UpperCamelCase* donde cada palabra comienza con una letra mayúscula: `NbrRep`

Todas estas formas de nombrar un objeto son equivalentes. En este libro usaremos el estilo *lowerCamelCase*. En general, debemos evitar los nombres que son demasiado largos, como `miNumeroDeRepeticionesDeMiExperimento` o demasiado cortos como `nR`, y los nombres que no permiten identificar los contenidos como `miVariable` o `miNumero`, asi que nombres como `a` o `b`. El objetivo es de tener una idea de lo que hay en cada objeto en base a su nombre.

Hay diferentes maneras de definir un nombre para los objetos que crearemos con R. En este libro, utilizamos el estilo *lowerCamelCase*. Lo importante no es la elección del estilo, sino la consistencia en su elección. El objetivo es tener un código funcional, pero también un código que sea fácil y agradable de leer para nosotros y para los demas.

Ahora que hemos elegido un nombre para nuestro objeto, debemos crearlo y hacer que R entienda que nuestro objeto debe contener el número 5. Hay tres maneras de crear un objeto bajo R:

- con <-
- con =
- o con ->

```
nbrRep <- 5
nbrRep = 5
5 -> nbrRep
```

En este libro siempre usaremos la forma <- para coherencia y también porque es la forma más común.

```
nbrRep <- 5
```

Acabamos de crear un objeto `nbrRep` y establecerlo con el valor *5*. Este objeto ahora está disponible en nuestro entorno de computación y puede ser utilizado. Algunos ejemplos :

```
nbrRep + 2
```

```
## [1] 7
```

```
nbrRep * 5 - 45/56
```

```
## [1] 24.19643
```

```
pi * nbrRep^2
```

```
## [1] 78.53982
```

El valor asociado con nuestro objeto `nbrRep` se puede modificar de la misma manera que cuando se creó:

```
nbrRep <- 5
nbrRep + 2
```

```
## [1] 7
```

```
nbrRep <- 10
nbrRep + 2
```

```
## [1] 12
```

```
nbrRep <- 5 * 2 + 7/3
nbrRep + 2
```

```
## [1] 14.33333
```

El uso de objetos tiene sentido cuando tenemos operaciones complejas para realizar y hace que el código sea más agradable de leer y entender.

```
(5 + 9^2 - 1/18) / (32 * 45/8 + 3)
```

```
## [1] 0.4696418
```

```
termino01 <- 5 + 9^2 - 1/18
termino02 <- 32 * 45/8 + 3
termino01 / termino02
```

```
## [1] 0.4696418
```

4.4 Los scripts

R es un lenguaje de programación denominado *lenguaje de scripting*. Esto se refiere al hecho de que la mayoría de los usuarios escribirán pequeñas piezas de código en lugar de programas completos. R se puede usar como una simple calculadora, y en este caso no será necesario mantener un historial de las operaciones que se han realizado. Pero si las operaciones a implementar son largas y complejas, puede ser necesario e interesante guardar lo que se ha hecho para poder continuar más adelante. El archivo en el que se almacenarán las operaciones es lo que comúnmente se llama el *script*. Un *script* es, por lo tanto, un archivo que contiene una sucesión de información comprensible por R y que es posible ejecutar.

4.4.1 Crear un script y documentarlo

Para abrir un nuevo script, es suficiente crear un archivo de texto vacío que será editado por un editor de texto como el bloc de notas en Windows o Mac OS, o Gedit o incluso nano en Linux. Por convención, este archivo toma la extensión ".r" o ".R" (lo mas comun). Esta última convención se usará en este libro (*"miArchivo.R"*). Desde la interfaz gráfica de R, es posible crear un nuevo script en Mac OS y Windows a través de *file*, luego *new script* y *save as*. Al igual que el nombre de los objetos, el nombre del script es importante para que podamos identificar fácilmente su contenido. Por ejemplo, podríamos crear un archivo `formRConceptsBase.R` que contenga los objetos que acabamos de crear y los cálculos que hicimos. Pero incluso con nombres de objetos y archivos bien definidos, será difícil recordar el significado de este archivo sin la documentación que acompaña a este script. Para documentar un script utilizaremos *comentarios*. Los *comentarios* son elementos que R identificará como tales y no se ejecutarán. Para especificar a R que vamos a hacer un *comentario*, debemos usar el carácter octothorpe (corsé o numeral) #. Los comentarios se pueden insertar en una nueva línea o al final de la línea.

```
# creación objeto número de repeticiones
nbrRep <- 5 # Comentario de fin de línea
```

Todo lo que hay despues del simbolo numeral # no sera ejecutado. Significa que podriamos usar comentarios como ### o `#comentario`, aun que se recomienda hacer comentarios con un solo simbolo numeral seguido por un espacio y despues su comentario: `# mi comentario`.

Los comentarios también se pueden usar para hacer que una línea ya no se ejecute. En este caso no queremos ejecutar la secunda linea:

```
nbrRep <- 5
# nbrRep + 5
```

Para volver a la documentación del script, se recomienda comenzar cada uno de nuestros scripts con una breve descripción de su contenido, luego cuando el script sea extenso, estructurarlo en diferentes partes para facilitar su lectura.

```
# ------------------------------------------------------------
# Aquí hay un script para adquirir los conceptos básicos
# con R
# fecha de creación : 27/06/2018
# autor : François Rebaudo
# ------------------------------------------------------------

# [1] creación del objeto número de repeticiones
# ------------------------------------------------------------

nbrRep <- 5

# [2] cálculos simples
# ------------------------------------------------------------

pi * nbrRep^2
```

```
## [1] 78.53982
```

Para ir más allá en el estilo del código, una guía completa de recomendaciones está disponible en línea en el sitio web *tidyverse* (en ingles ; `http://style.tidyverse.org/`).

4.4.2 Ejecutar un script

Como tenemos un script, no trabajamos directamente en la consola. Pero solo la consola puede *entender* el código R y devolvernos los resultados que queremos obtener. Por ahora, la técnica más simple es copiar y pegar las líneas que queremos ejecutar desde nuestro script hasta la consola. A partir de ahora, ya no utilizaremos editores de texto como bloc de notas, sino editores especializados para la creación de scripts R. Sera es el objetivo del siguiente capítulo.

4.5 Conclusión

Felicitaciones, hemos llegado al final de este primer capítulo sobre la base de R. Sabemos:

- Instalar R
- Usar R como una calculadora
- Crear **objetos** y utilisarlos para los calculos aritméticos
- Elejir nombres pertinentes para los objetos
- Crear un nuevo **script**
- Elejir un nombre pertinente para el archivo del script
- Ejecutar el codigo de un script
- Documentar los scripts con **comentarios**
- Usar un estilo de código para que sea agradable de leer y facil de entender

Chapter 5

Elejir un entorno de desarrollo

5.1 Editores de texto y entorno de desarrollo

Hay muchos editores de texto, el capítulo anterior permitió introducir algunos de los más simples como el bloc de notas de Windows. Rápidamente los límites de estos editores han hecho que la tarea de escribir un script tedioso. De hecho, incluso estructurando su script con comentarios, sigue siendo difícil ubicarse en este. Aquí es donde entran los editores de texto especializados para facilitar la escritura y la lectura de scripts. El editor de texto para R más común es Rstudio, pero hay muchos más. Hacer una lista exhaustiva de todas las soluciones disponibles está más allá del alcance de este libro, por lo que nos centraremos en las tres soluciones que utilizo a diario: **Notepad ++**, **Rstudio** y **Geany**.

5.2 RStudio

5.2.1 Instalar RStudio

El programa para instalar el software RStudio se encuentra en la parte *Products* del sitio web de RStudio (`https://www.rstudio.com/`). Instalaremos RStudio para uso local (en nuestra computadora), por lo que la versión que nos interesa es *Desktop*. Usaremos la versión *Open Source* que es gratis. Luego, seleccionamos la versión que corresponde a nuestro sistema operativo (Windows, Mac OS, Linux), descargamos el archivo correspondiente y lo ejecutamos para comenzar la instalación. Podemos mantener las opciones predeterminadas durante la instalación.

5.2.2 Un script con RStudio

Podemos abrir RStudio. En la primera apertura, la interfaz se divide en dos con la consola R a la izquierda que vimos en el capítulo anterior (Figura 5.2). Para abrir un nuevo script, vamos al menú *Archivo* (o *File*), *Nuevo archivo* (o *New File*), *R script*. Por defecto, este archivo tiene el nombre *Untitled1*. Hemos visto en el capítulo anterior la importancia de dar un nombre pertinente a nuestros scripts, por lo que lo cambiaremos de nombre a *selecEnvDev.R*, en el menú *Archivo* (o *File*), con la opción *Guardar como …* (o *Save As…*). Podríamos notar que el lado izquierdo de RStudio ahora está dividido en dos, con la consola en la parte inferior de la pantalla y el script en la parte superior.

Figure 5.1: Logo RStudio.

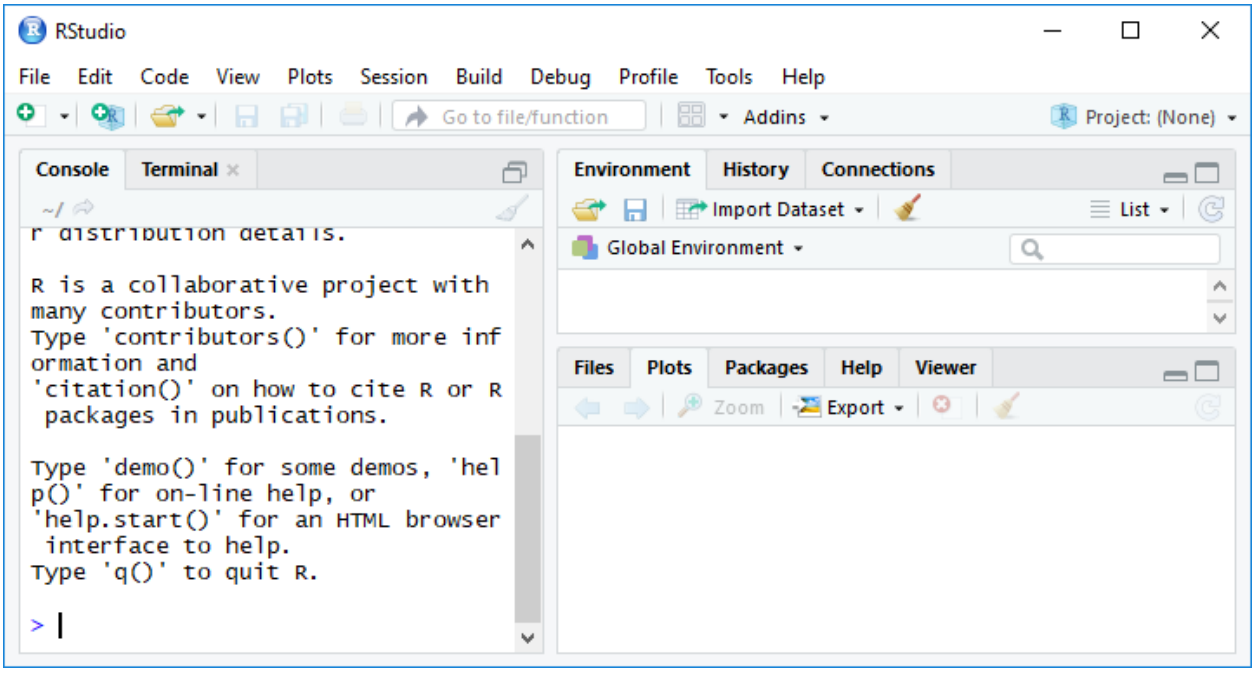

Figure 5.2: Captura de pantalla de RStudio en Windows: pantalla por defecto.

Luego podemos comenzar a escribir nuestro script con los comentarios que describan lo que vamos a encontrar allí, y agregar un cálculo simple. Una vez que hayamos copiado el siguiente código, podemos guardar nuestro script con el comando CTRL + S o yendo a *Archivo* (o *File*, luego *Guardar* (o *Save*).

```
# ------------------------------------------------------------
# Un script para seleccionar su entorno de desarrollo
# fecha de creación : 27/06/2018
# autor : François Rebaudo
# ------------------------------------------------------------

# [1] cálculos simples
# ------------------------------------------------------------
nbrRep <- 5
pi * nbrRep^2
```

```
## [1] 78.53982
```

Para ejecutar nuestro script, simplemente seleccionamos las líneas que deseamos ejecutar y usamos la combinación de teclas CTRL + ENTER. El resultado aparece en la consola (Figura 5.3).

Podemos ver que, de forma predeterminada, en la parte del script, los comentarios aparecen en verde, los números en azul y el resto del código en negro. En la parte de la consola, lo que se ejecutó aparece en azul y los resultados de la ejecución en negro. También podemos observar que en la parte del código cada línea tiene un número correspondiente al número de línea a la izquierda sobre un fondo gris. Este es el resaltado de preferencias de sintaxis predeterminado con RStudio. Estas preferencias de sintaxis pueden modificarse yendo al menú *Herramientas* (o *Tools*), *Opciones globales …* (o *Global Options…*), *Aspecto* (o *Appearance*), y luego seleccionando otro tema del *Editor de tema:* (o *Editor theme:*). Elegiremos el tema *Cobalt*, luego *OK* (Figura 5.4).

Sabemos cómo crear un nuevo script, guardarlo, ejecutar su contenido y cambiar la apariencia de RStudio. Veremos los muchos otros beneficios de RStudio a lo largo de este libro, ya que es el entorno de desarrollo que se utilizará. Sin embargo, seremos

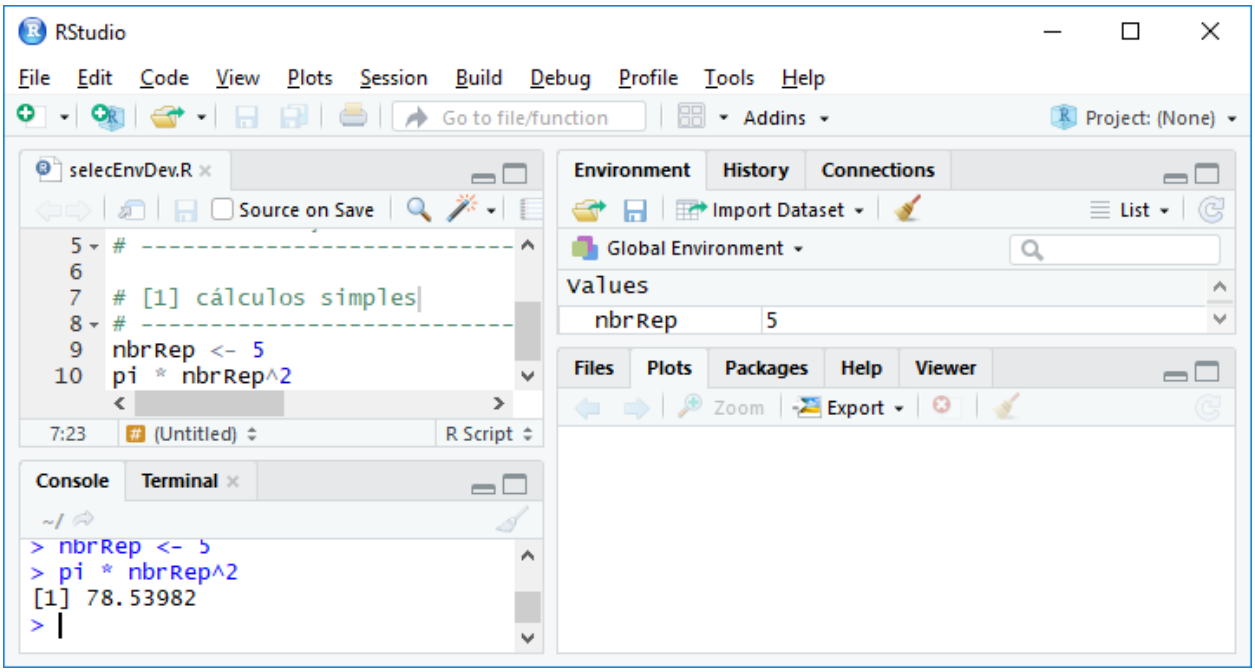

Figure 5.3: Captura de pantalla de RStudio en Windows: ejecutar nuestro script con CTRL + ENTER.

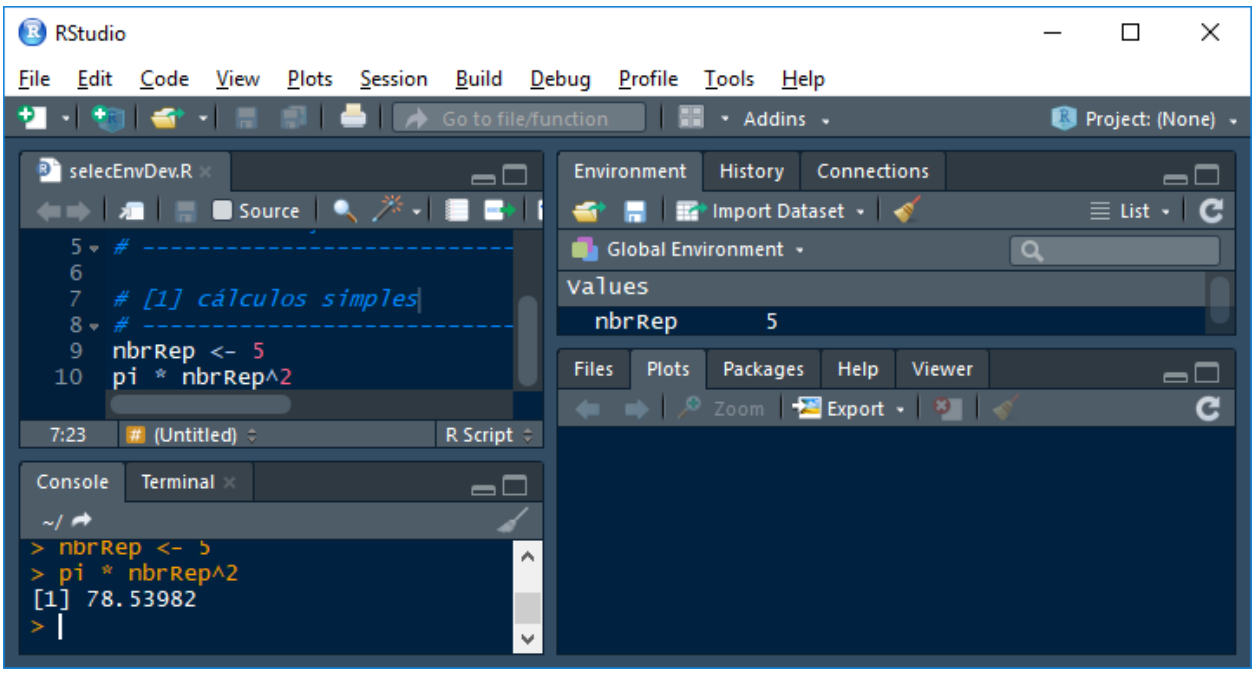

Figure 5.4: Captura de pantalla de RStudio en Windows: cambiar preferencias de sintaxis.

Figure 5.5: Logo Notepad++

especialmente cuidadosos de que todos los scripts desarrollados a lo largo de este libro se ejecuten de la misma manera, independientemente del entorno de desarrollo utilizado.

5.3 Notepad++ avec Npp2R

5.3.1 Installer Notepad++ (pour Windows uniquement)

Le programme pour installer Notepad++ se trouve dans l'onglet *Downloads* (`https://notepad-plus-plus.org/download/`). Vous pouvez choisir entre la version 32-bit et 64-bit (64-bit si vous ne savez pas quelle version choisir). Notepad++ seul est suffisant pour écrire un script, mais il est encore plus puissant avec *Notepad to R* (*Npp2R*) qui permet d'exécuter automatiquement nos script dans une console en local sur notre ordinateur ou à distance sur un serveur.

5.3.2 Installer Npp2R

Le programme pour installer Npp2R est hébergé sur le site de Sourceforge (`https://sourceforge.net/projects/npptor/`). Npp2R doit être installé après Notepad++.

5.3.3 Un script avec Notepad++

Lors de la première ouverture Notepad++ affiche un fichier vide *new 1* (Figura 5.6).

Puisque nous avons déjà créer un script pour le tester avec RStudio, nous allons l'ouvrir à nouveau avec Notepad++. Dans *Fichier*, selectionnons *Ouvrir…* puis choisir le script *selecEnvDev.R* créé précédemment. Une fois le script ouvert, allons dans *Langage*, puis *R*, et encore une fois *R*. La coloration syntaxique apparaît (Figura 5.7).

L'execution du script ne peut se faire que si Npp2R est en cours d'exécution. Pour se faire il est nécessaire de lancer le programme Npp2R depuis l'invite de Windows. Un icône devrait apparaître en bas de votre écran. L'exécution automatique du code depuis Notepad++ se fait en sélectionnant le code à exécuter puis en utilisant la commande `F8`. Si la commande ne fonctionne pas et que vous venez d'installer Notepad++, il est peut être nécessaire de redémarrer votre ordinateur. Si la commande fonctionne, une nouvelle fenêtre va s'ouvrir avec une consol exécutant les lignes souhaitées (Figura 5.8.

Comme pour RStudio, la coloration syntaxique peut être modifiée depuis le menu *Paramètres*, et un nouveau thème peut être sélectionné (par exemple *Solarized* dans la Figura 5.9)

Par rapport aux autres éditeurs de texte, Notepad++ a l'avantage d'être très léger et offre une vaste gamme d'options pour personnaliser l'écriture du code.

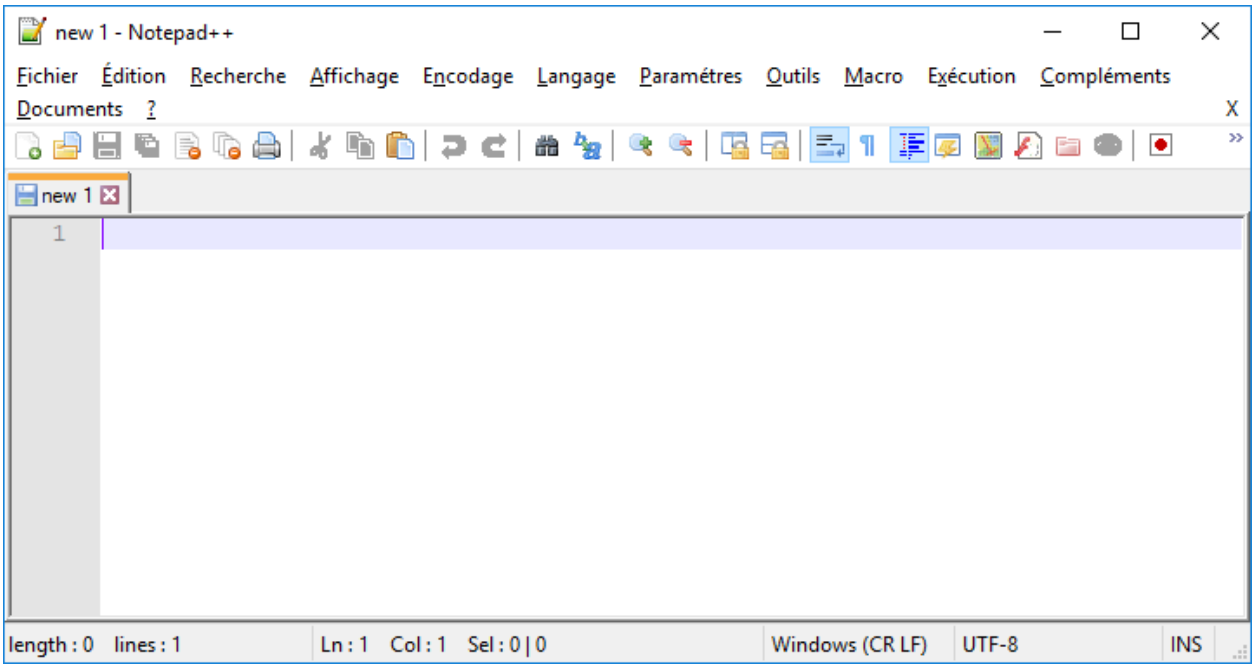

Figure 5.6: Captura de pantalla de Notepad++ en Windows: pantalla por defecto.

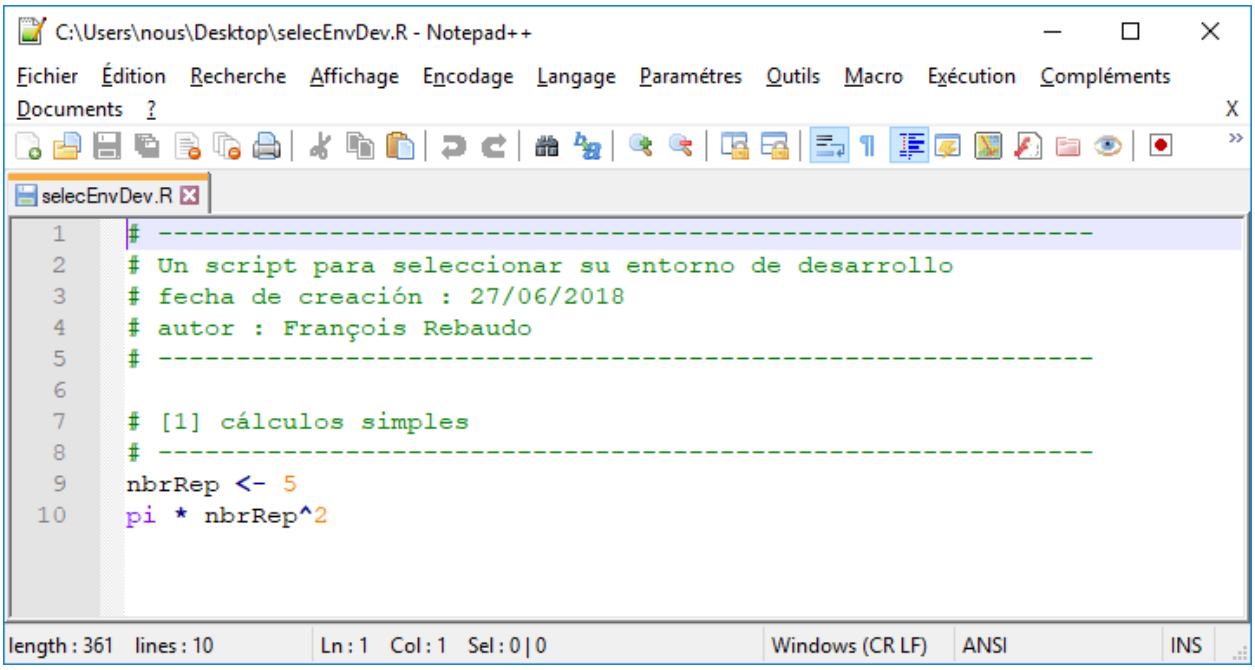

Figure 5.7: Captura de pantalla de Notepad++ en Windows: ejecutar nuestro script con F8.

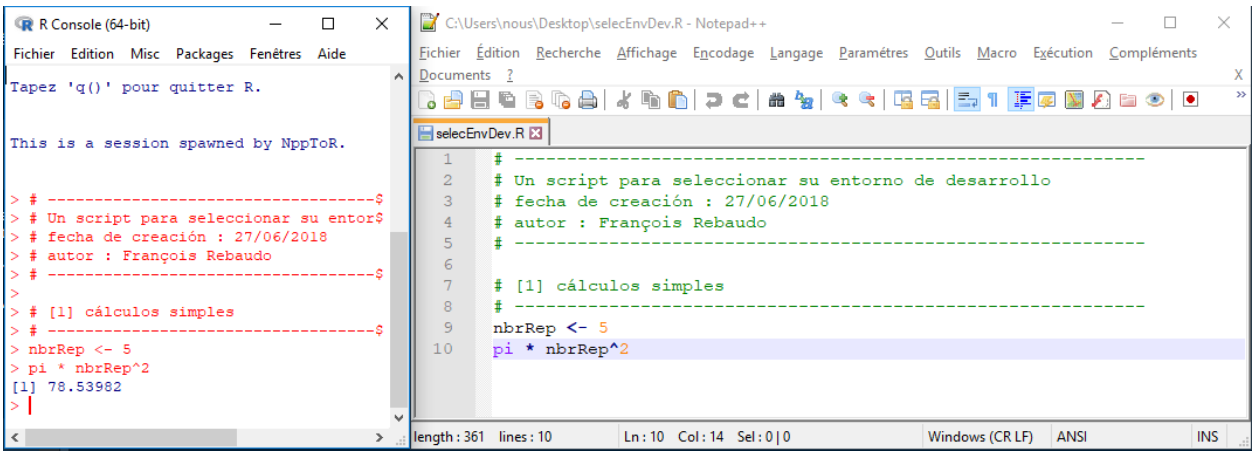

Figure 5.8: Captura de pantalla de Notepad++ en Windows: la consola con F8.

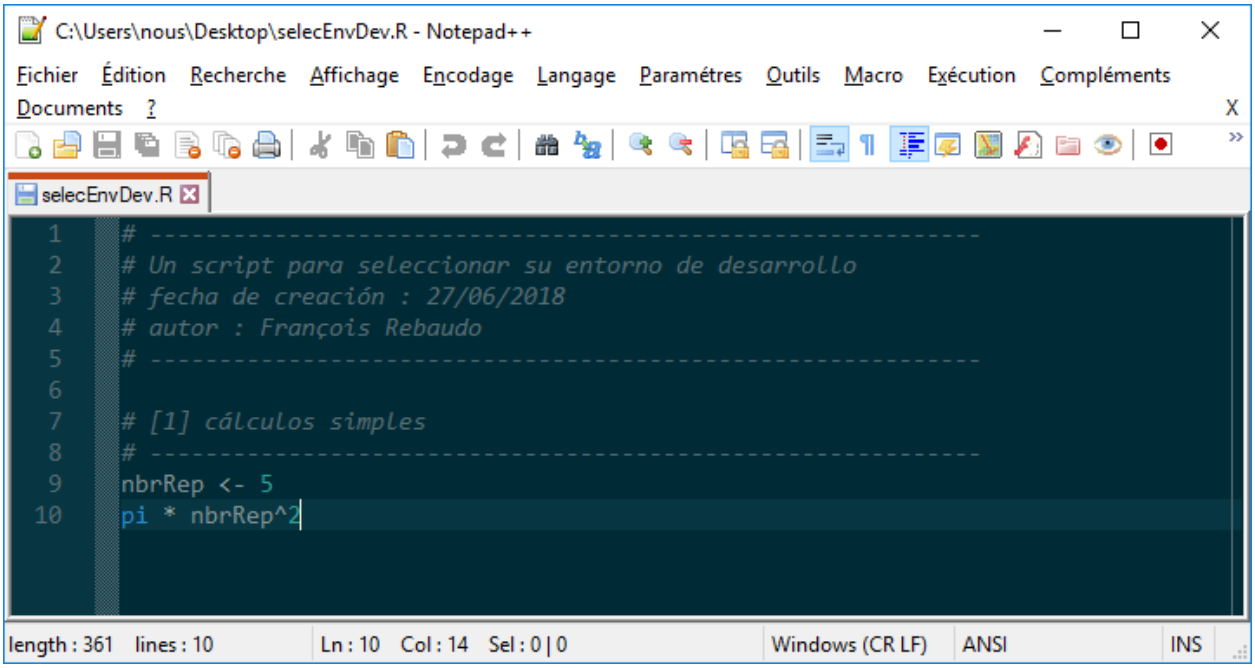

Figure 5.9: Captura de pantalla de Notepad++ en Windows con el tema Solarized.

Figure 5.10: Logo Geany

5.4 Geany (pour Linux et Windows)

5.4.1 Installer Geany

XXX `https://www.geany.org/`

5.4.2 Un script avec Geany

XXX

5.5 Autres solutions

Il existe beaucoup d'autres solutions, certaines spécialisées pour R comme **Tinn-R** (`https://sourceforge.net/projects/tinn-r/`), et d'autres plus généralistes pour la programmation comme **Atom** (`https://atom.io/`), **Sublime Text** (`https://www.sublimetext.com/`), **Vim** (`https://www.vim.org/`), **Gedit** (`https://wiki.gnome.org/Apps/Gedit`), **GNU Emacs** (`https://www.gnu.org/software/emacs/`), ou encore **Brackets** (`http://brackets.io/`) et **Eclipse** (`http://www.eclipse.org/`).