

CAPÍTULO 1

Introducción al ambiente R

Jéssica Karina da Silva Pachú, José Bruno Malaquias, Tatiane Caroline Grella, Milton Fernando Cabezas Guerrero

Resumen

R cuenta con una gama de herramientas como pruebas paramétricas y no paramétricas, análisis de regresión lineal y no lineal, análisis de supervivencia, análisis multivariado, producción gráfica, entre otros. Es importante mencionar que R no es simplemente un programa estadístico. En este capítulo, presentamos aspectos introductorios al entorno R, como instalar el programa, cargar e instalar paquetes. Los comandos a ejecutar en R se muestran en color azul.

Palabras claves: lenguaje R; presentación; instalación; cargando; paquetes.

1. Introducción

La estadística se puede definir como matemáticas aplicadas a datos de observación [1]. A través de la Estadísticas obtenemos descripciones claras, sintéticas y objetivas [2].

Desde una perspectiva entomológica, la estadística es una condición fundamental ya que trabaja con métodos científicos para recolectar, organizar, resumir y presentar datos, además, de obtener conclusiones y tomar decisiones.

Las dificultades a las que se enfrentan los estudiantes e investigadores entomológicos que necesitan hacer uso de la inferencia estadística parece ser el resultado de la falta de libros didácticos escritos en un lenguaje adecuado para ellos, libres de los algebraísmos que suelen emplear los estadísticos matemáticos. Para cubrir esa brecha, este libro tiene como objetivo dar a conocer los comandos básicos para el análisis descriptivo, análisis de varianza y análisis de desviación, con datos entomológicos en lenguaje R. La estrategia empleada

para lograr este propósito consistió en preparar *scripts* en un lenguaje sencillo y con ejemplos entomológicos.

El programa estadístico que ha sido más útil en el análisis estadístico de datos biológicos es el basado en el entorno R, que es un lenguaje de programación estadística de acceso abierto que se ha vuelto cada vez más popular.

Por sus características, el *R Development Core Team* lo clasifica como un entorno R [3].

R es un software gratuito para el análisis de datos. Esto significa que se puede utilizar, copiar, distribuir, modificar y mejorar libremente. Fue desarrollado en 1996, por los profesores de estadística Ross Ihaka y Robert Gentleman, que trabajaron y fueron colegas en el departamento de estadística de la Universidad de Auckland, los dos compartieron un interés en la estadística computacional y vieron la necesidad de un mejor entorno de software para el área. Así que desarrollaron un nuevo lenguaje computacional, similar al lenguaje S [3].

El software es un entorno de investigación y desarrollo colaborativo internacional formalmente mantenido por la Fundación R, que ofrece una amplia gama de técnicas estadísticas y gráficas, que incluyen modelado lineal y no lineal, pruebas estadísticas clásicas, análisis de series de tiempo, clasificación, agrupamiento y otros. Puede "descargarlo" de Internet, sin costo (www.r-project.org). Un conjunto básico de paquetes viene integrado en la instalación de R, con muchos más disponibles en la red de distribución de R (CRAN). Investigaciones realizadas con profesionales en el área que la popularidad de R ha aumentado sustancialmente en los últimos años [4]. A partir de 2017, R ya tiene más de 10000 paquetes disponibles [5]. Un paquete es un conjunto de funciones a las que se puede acceder a su código simplemente importándolo. Hay paquetes para resolver diversos problemas como gráficos, análisis descriptivo, análisis de varianza, etc.

2. Motivos para el uso de R

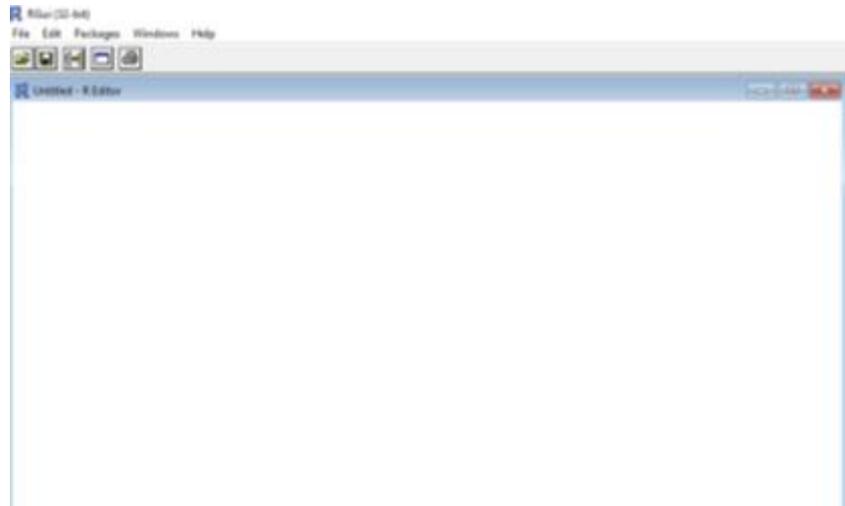
- Todos los códigos son abiertos, reproducibles y adaptables;
- Compatible con Windows, Mac, Linux y etc;
- Se utiliza para crear otro software;
- Las empresas públicas y privadas están prestando atención a R;

- Enlace con C, C ++, Java, entre otros;
- Gran documentación (cambios de versión y ayuda);
- Comunidad académica difundiendo conocimientos;
- Altamente extensible;
- La comunidad lo apoya continuamente, con correcciones rápidas de errores y nuevas funciones.

3. Instalación de R

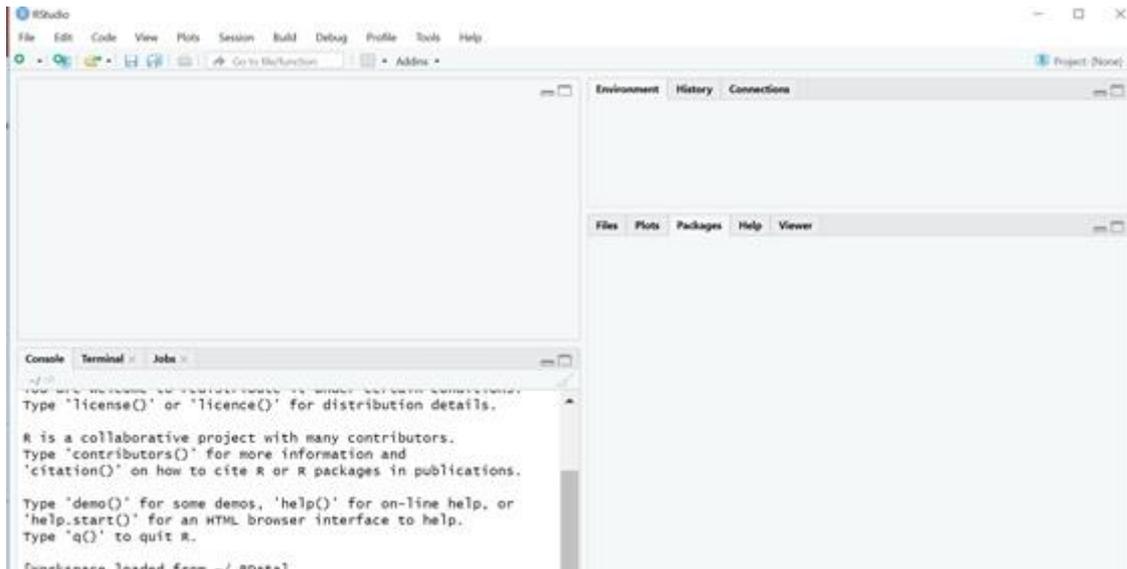
- Haga clic en CRAN;
- Elija el espejo de su elección (CRAN mirrors);
- Haga clic en Windows;
- Haga clic en la base y guarde el archivo R para Windows;
- Entonces simplemente ejecute el archivo.

4. Consola de software "R"



5. Uso de R como un entorno de desarrollo integrado (IDE)

RStudio es una interfaz integrada para gráficos y cálculos estadísticos. Accesorio escrito en lenguaje C++, está disponible en dos ediciones: RStudio Desktop, que se ejecuta localmente como una aplicación de escritorio estándar; y RStudio Server, que le permite acceder a RStudio mediante un navegador web.



Para **bajar RStudio** haga clic en: <https://www.rstudio.com/products/rstudio/download/> y escoja la opción *Rstudio desktop*

6. Script

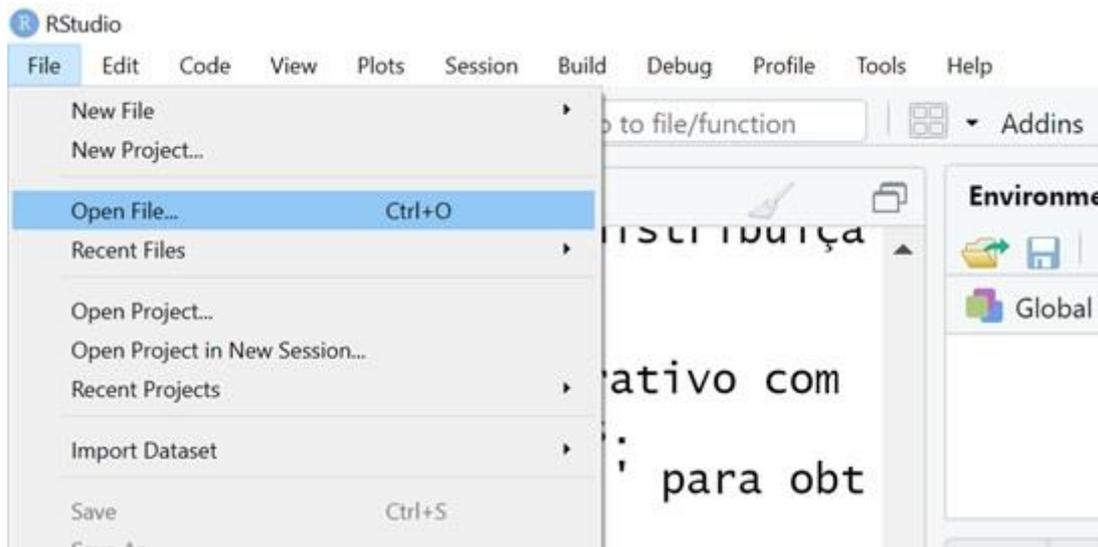
Para abrir y practicar en Rstudio con un nuevo script, vaya a:

- File > New File > R Script

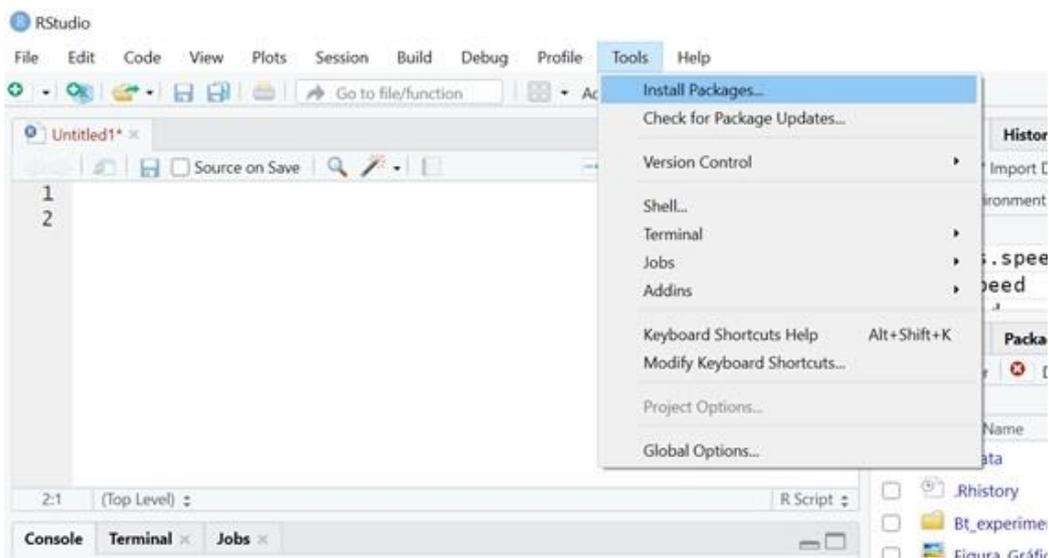
Si prefiere abrir un script que ya ha guardado, vaya a:

- Open File > seleccione el archivo > Abrir

En el script puede escribir comandos a ejecutar y también comentarios. Los comandos se escriben como en la consola y todo lo que se escribe después del # se considera solo como comentarios.

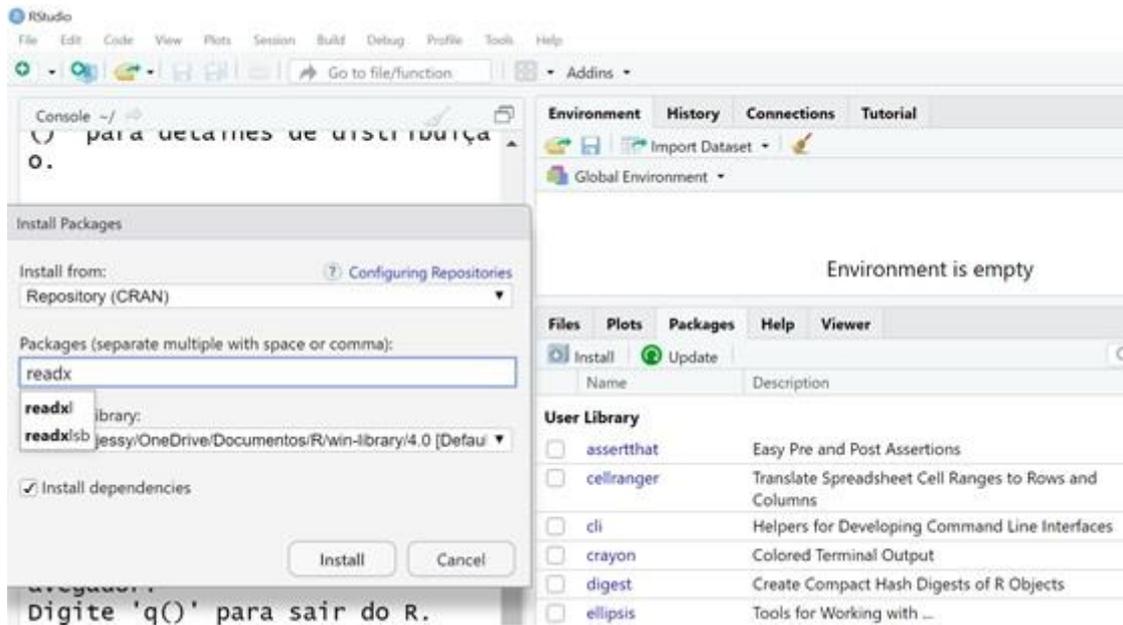


R es un programa que ocupa poco espacio y memoria y generalmente se ejecuta rápido, eso se debe a que los paquetes vienen en la instalación base, es decir, configuraciones mínimas para funcionar. Para realizar tareas más complejas puede ser necesario instalar paquetes adicionales (packages), este es uno de los grandes beneficios de R, su gran colección de paquetes. Cualquiera puede enviar un paquete y cada código enviado está disponible en Internet. Sin embargo, existe un proceso de evaluación por el que pasa el código y se deben cumplir ciertas reglas estrictas con respecto al formato del código, el manual de usuario y la forma de actualizar el paquete.



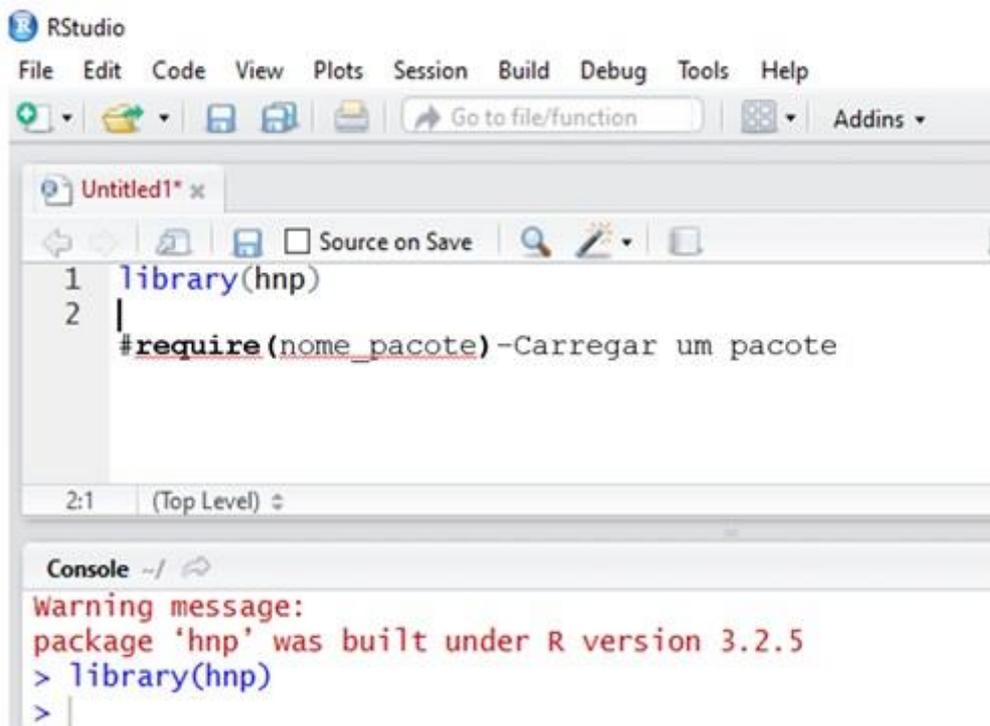
Otra forma es, en el lado derecho de la pantalla, hacer clic en:

- Packages>Install> ingrese el nombre del paquete e instálelo.



7. Cargando un paquete

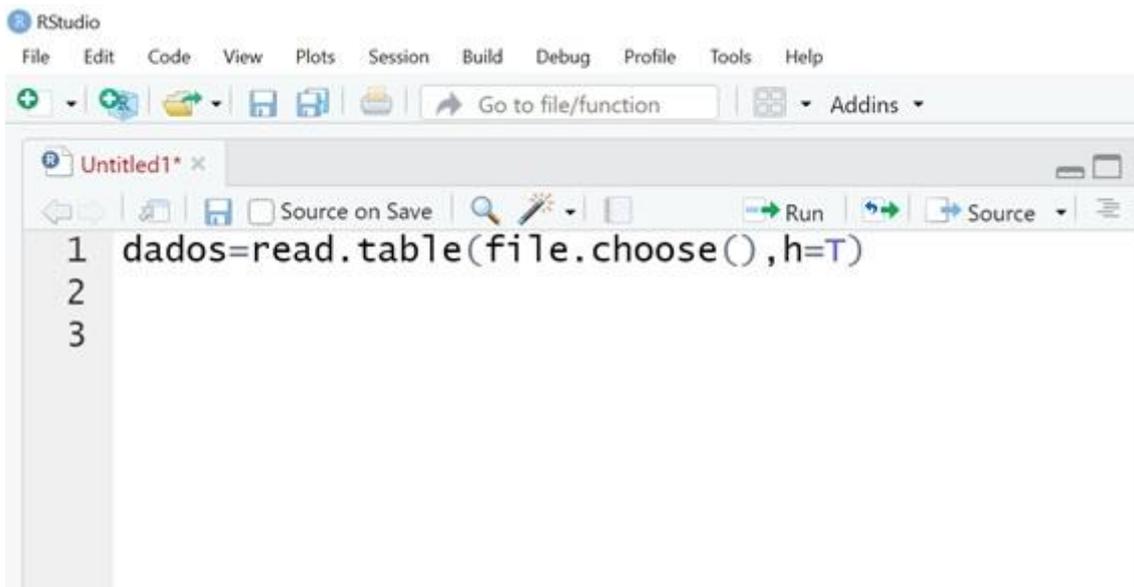
Para cargar un paquete dentro de un script, usamos la función `library()` o `require()`, dentro del paréntesis el nombre del paquete que se debe cargar.



8. Leer un archivo

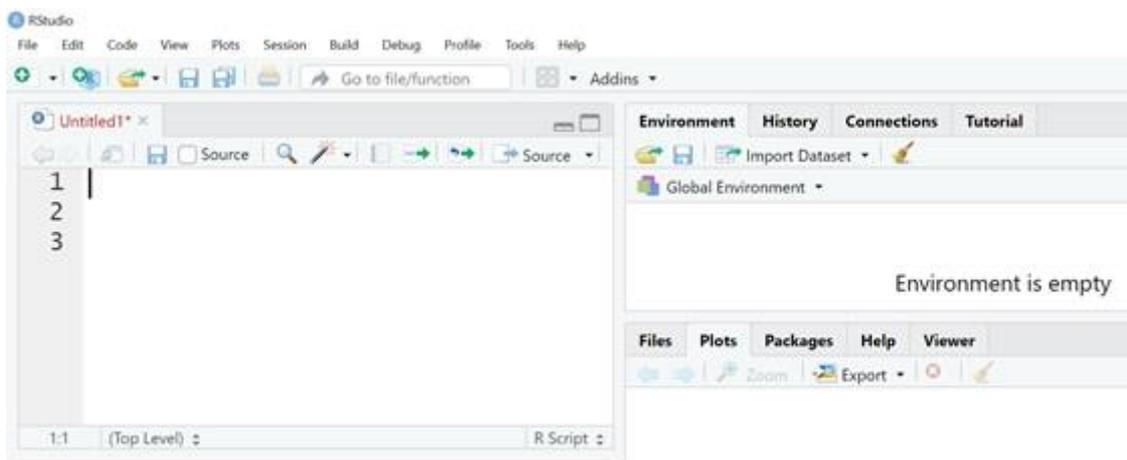
Para leer una base de datos, use la función `read.table` para importar los datos.

La línea de comando ejemplificada `datos=read.table(file.choose(),h=T)` lo dirige a la carpeta que contiene su base de datos en el entorno R (datos), la función para importar los datos (`read.table`) y la función para elegir la carpeta donde están sus datos (`file.choose`).



9. Visualización de un Plot

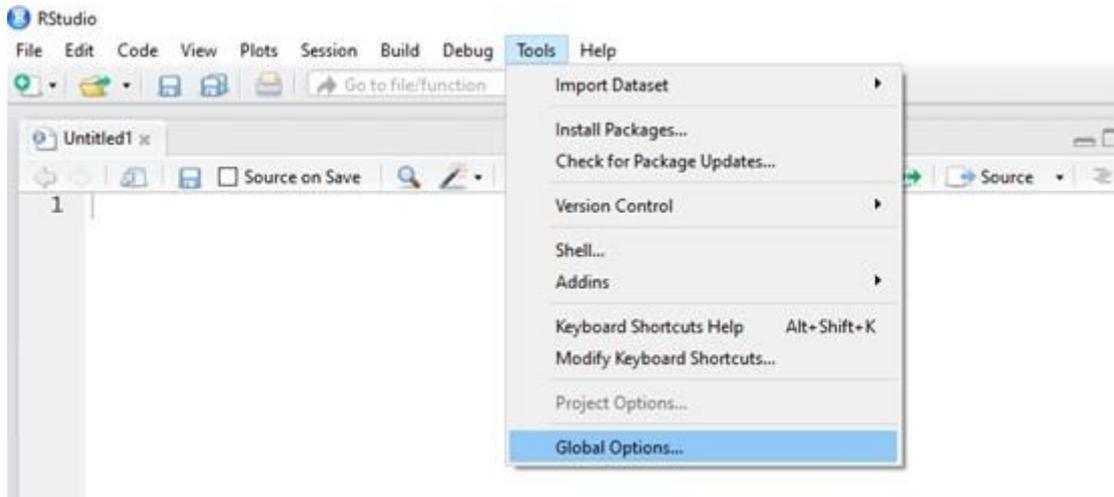
Para ver sus gráficos, haga clic en la pestaña plots en la fila superior de la consola o en la pestaña plots en el lado derecho, de esta manera puede usar la escoba para limpiar o hacer zoom para ampliar su plot.



10. Personaliza tu RStudio

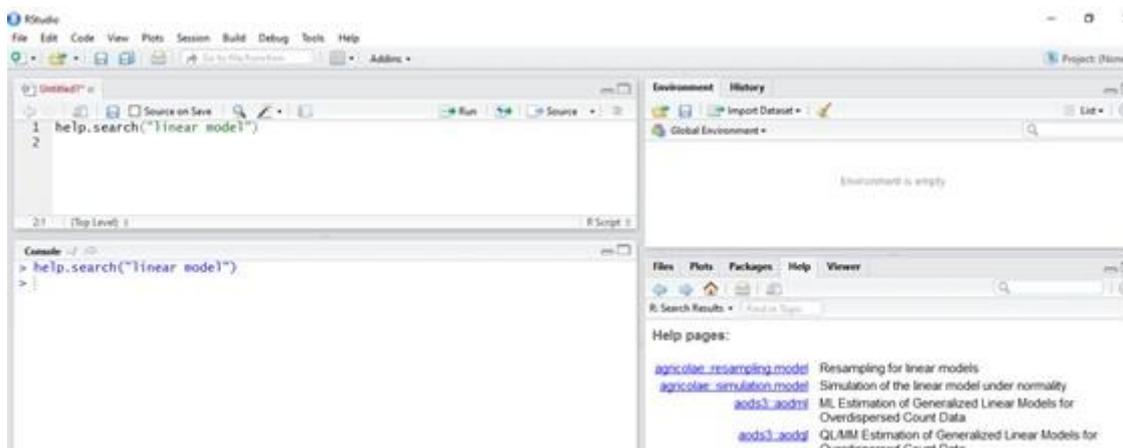
En esta opción puede cambiar y personalizar RStudio según sus preferencias, por ejemplo, modificar la fuente, tamaño, color, etc. Haga clic en:

- Tools> Global Options>



11. Buscando ayuda

Existen algunas formas de obtener ayuda para ejecutar comandos y paquetes. Por ejemplo, haciendo clic en la pestaña Help o escribiendo la línea de comando `help.search ()`



12. Algunas funciones y comandos básicos

Para ver la base de datos

View ()

Limpiar la consola

Ctrl+L

Para leer las líneas de comando

Ctrl+Enter ou Ctrl+R

Cambio de directorio

Ctrl+Shift+H:

getwd()

Para ver los nombres de los archivos dentro de la carpeta elegida para trabajar:

list.files()

13. Referencias de los paquetes utilizados

- [1] Pimentel-Gomes F. Curso de estatística experimental. 15. Ed. Piracicaba: FEALQ, 2009.
- [2] Nazareth H. Curso básico de estatística. Ática, São Paulo, 2003.
- [3] Peternelli L.A, Mello M.P. Conhecendo o R: uma visão estatística. Viçosa: UFV, v. 1, 2011.
- [4] Muenchen R.A. The popularity of data analysis software. URL <http://r4stats.com/popularity>, 2012.
- [5] Smith D. CRAN now has 10,000 R packages. Here's how to find the ones you need, 2017.

14. Referencias recomendadas

- Borcard D., Gillet F., Legendre P. Numerical ecology with R. Springer, 2018.
- British Ecological Society. A guide to data management in ecology and evolution. 2014.
- Crawley M.J. The R book. John Wiley & Sons, 2012.
- Matloff Norman. The art of R programming: A tour of statistical software design. No Starch Press, 2011.
- Peternelli L.A., Mello M.P. Conhecendo o R: uma visão estatística. Viçosa: UFV, v. 1, 2011.