



Introducción



¿Qué es R?

- Conjunto de programas para la manipulación de datos, cálculo y gráficos.
- Cuenta con: almacenamiento de datos, operadores para realizar cálculos con variables, herramientas para análisis de datos, posibilidades gráficas para análisis, lenguaje de programación desarrollado.



Ventajas

- 100% gratuito, comunidad amplia de programadores.
- Versatilidad: nos permite hacer desde cálculos sencillos de estadística descriptiva, realizar simulaciones, graficar.
- Se puede combinar con otros softwares (Latex) para hacer documentos escritos.
- Análisis efectuados en R son transparentes, fáciles de compartir y replicar.

Recursos: cheatsheets

Cheatsheet	Autor/a	Link
R Basics	Tom Short	https://cran.r-project.org/doc/contrib/Short-refcard.pdf
Advanced R	Arianne Colton y Sean Chen	https://www.rstudio.com/wp-content/uploads/2016/02/advancedR.pdf
Base R	<u>Mhairi McNeill</u>	http://github.com/rstudio/cheatsheets/raw/master/base-r.pdf
Strings	Rstudio	https://github.com/rstudio/cheatsheets/raw/master/strings.pdf
Data import	Rstudio	https://github.com/rstudio/cheatsheets/raw/master/data-import.pdf
Data transformati on	Rstudio	https://github.com/rstudio/cheatsheets/raw/master/data-import.pdf
RStudio application	Rstudio	https://github.com/rstudio/cheatsheets/raw/master/rstudio-ide.pdf
Ggplot2	Rstudio	https://github.com/rstudio/cheatsheets/raw/master/data-visualization-2.1.pdf
RMarkdown	Rstudio	https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf

Más recursos: internet

- www.stackoverflow.com
- www.stackexchange.com
- www.r-bloggers.com
- blog.revolutionanalytics.com
- <https://www.rstudio.com/resources/cheatsheets/>
- <https://jrnold.github.io/qss-tidy/introduction.html>

Instalación

- R es el software que contiene todo el lenguaje de programación; Rstudio es una versión mucho más amigable.

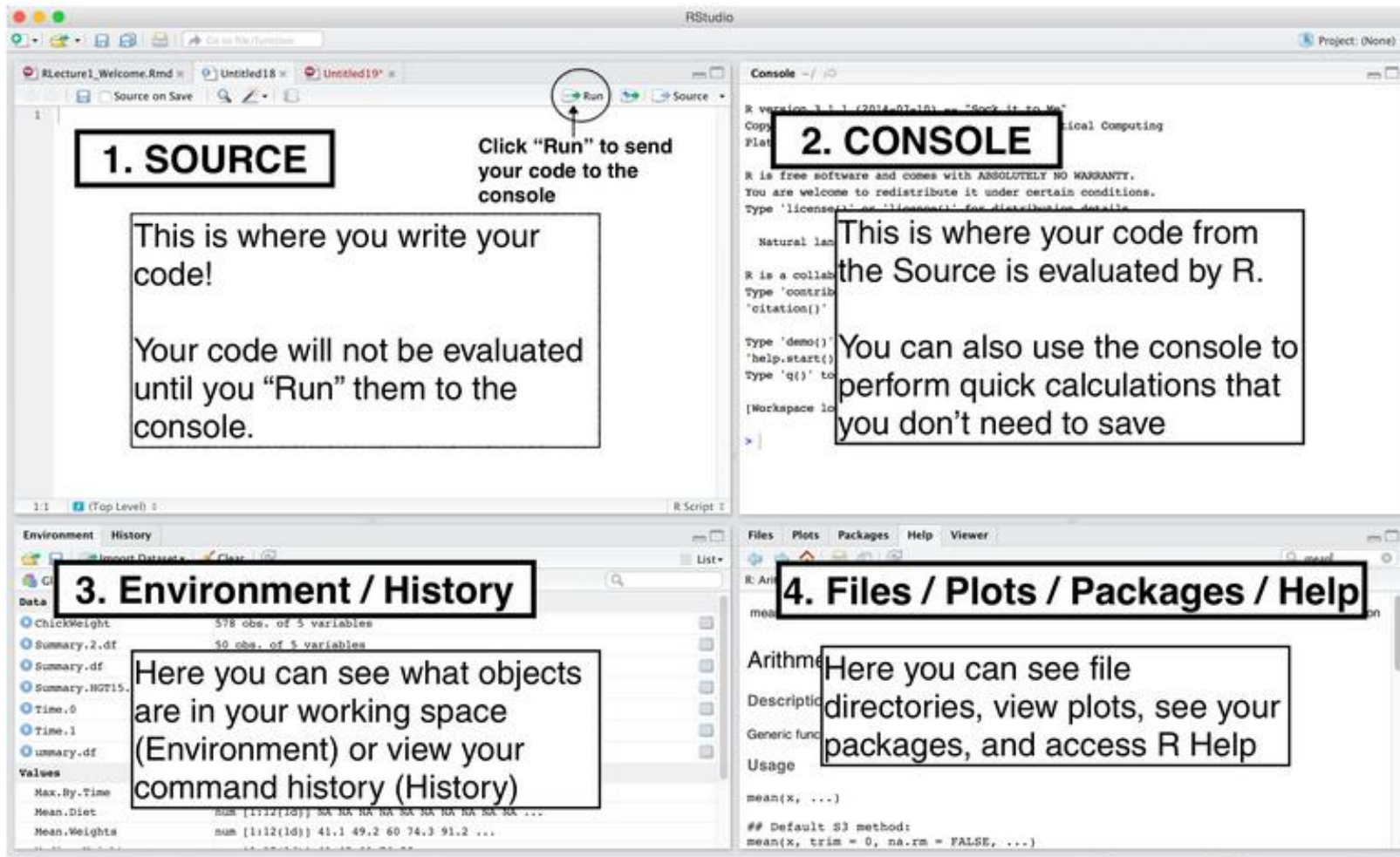
- Para descargar R

En Windows: <http://cran.r-project.org/bin/windows/base/>

En Mac: <http://cran.r-project.org/bin/macosx/>

- Para descargar Rstudio

<http://www.rstudio.com/products/rstudio/download/>



Ventanas

Paquetes

- Si queremos acceder a datos y códigos escritos por otras personas debemos instalar paquetes.
- R clasifica a los paquetes como base (default) o recomendados.
- Instalar un paquete significa que lo guardamos en la computadora. Normalmente lo instalamos de un repositorio que se llama CRAN.
- Cuando lo instalamos, ya no lo tenemos que instalar más. Sin embargo tenemos que “llamar” cada vez que lo usamos.

Vectores, escalares, objetos

- Vectores: una combinación de números o caracteres en una columna.
- Escalar: un valor
- Objeto: puede ser un número, una base de datos, un test estadístico o un resumen de un análisis. Pueden tener diversos atributos.
- El objeto se crea: `objeto <-`. Para verlos, los “imprimimos”.

Funciones y matrices

- Función: es un procedimiento que toma uno o dos objetos como argumentos, hace algo, y luego muestra un nuevo objeto.
- Podemos pensar en las funciones como procedimientos para decirle a R que haga diferentes cosas.
- Matriz: a diferencia de vectores contiene filas y columnas. Debe ser de números o caracteres, a diferencia de dataframe que combina ambos.

Funciones para crear vectores

Función	Ejemplo	Resultado
<code>c(a, b, ...)</code>	<code>c(1, 5, 9)</code>	1, 5, 9
<code>a:b</code>	<code>1:5</code>	1, 2, 3, 4, 5
<code>seq(from, to, by, length.out)</code>	<code>seq(from = 0, to = 6, by = 2)</code>	0, 2, 4, 6
<code>rep(x, times, each, length.out)</code>	<code>rep(c(7, 8), times = 2, each = 2)</code>	7, 7, 8, 8, 7, 7, 8, 8

Funciones para crear matrices

<code>cbind(a, b, c)</code>	Combine vectors as columns in a matrix	<code>cbind(1:5, 6:10, 11:15)</code>
<code>rbind(a, b, c)</code>	Combine vectors as rows in a matrix	<code>rbind(1:5, 6:10, 11:15)</code>
<code>matrix(x, nrow, ncol, byrow)</code>	Create a matrix from a vector x	<code>matrix(1:12, nrow = 3, ncol = 4)</code>
<code>data.frame()</code>	Create a dataframe from named columns	<code>data.frame("age" = c(19, 21), sex = c("m", "f"))</code>

Errores comunes

- Cuando está tratando de correr un código previo (+)
- Podemos escribir mal una función
- Puntuación

Importar y exportar bases de datos

- Para importar y exportar bases, lo importante es tener claro el directorio de trabajo.
- Es una carpeta donde automáticamente se guardan los archivos.

Filtrar por variables y por observaciones

- Podemos hacer un “subset” es decir, una base más pequeña con las variables que vamos a usar.
- También podemos filtrar la base utilizando observaciones que queremos (ejemplo: filtrar sólo por mujeres, o por cierto país/región, etc).

Manipulación de datos

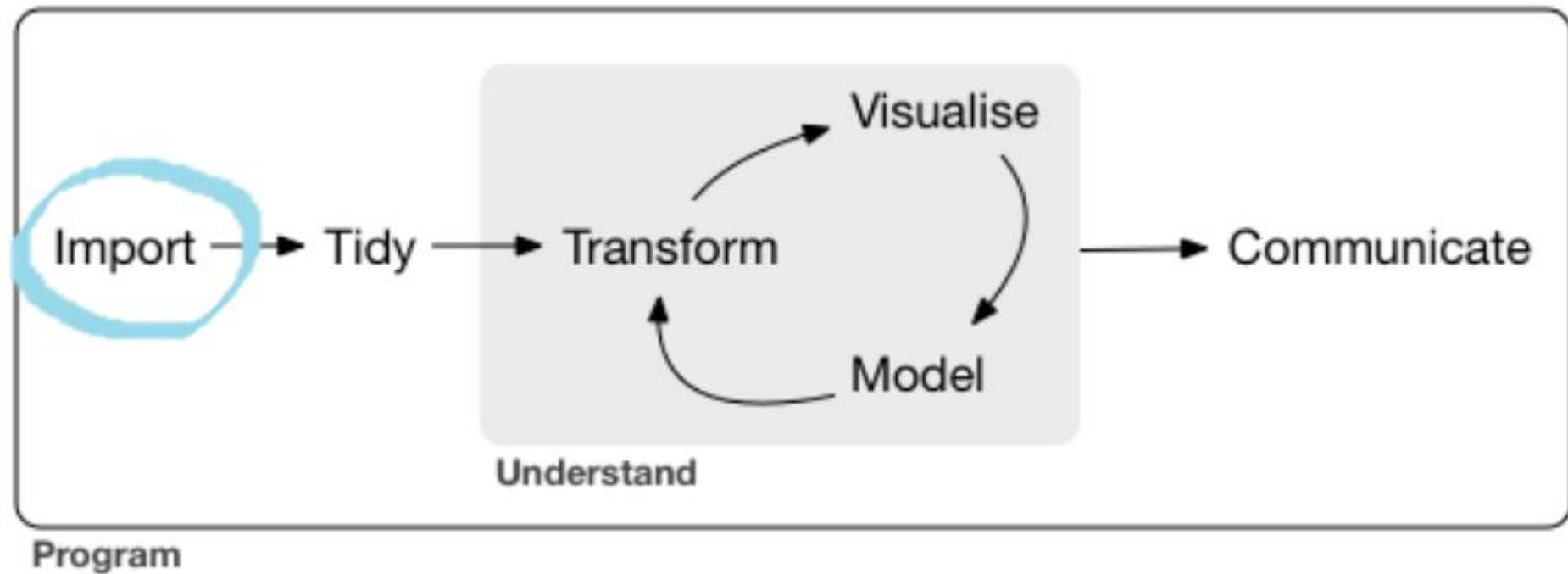


Figure 4.1: Primera etapa: Importar datos (<http://r4ds.had.co.nz/>)

Paquete *dplyr*

- Desarrollado por Hadley Wickham de Rstudio.
- Ventaja: proporciona una “gramática” para la manipulación y operaciones con bases de datos.
- Algunos de los principales “verbos” de dplyr son:

select: devuelve un conjunto de columnas

filter: devuelve un conjunto de filas según una o varias condiciones lógicas

arrange: reordena filas de un data frame

rename: renombra variables en una data frame

mutate: añade nuevas variables/columnas o transforma variables existentes

summarise/summarize: genera resúmenes estadísticos de diferentes variables en el data frame, posiblemente con *strata*

%>% : el operador "pipe" es usado para conectar múltiples acciones en una única "pipeline" (tubería)

Paquete *dplyr*

- Desarrollado por Hadley Wickham de Rstudio.
- Ventaja: proporciona una “gramática” para la manipulación y operaciones con bases de datos.
- Algunos de los principales “verbos” de dplyr son:

select: devuelve un conjunto de columnas

filter: devuelve un conjunto de filas según una o varias condiciones lógicas

arrange: reordena filas de un data frame

rename: renombra variables en una data frame

mutate: añade nuevas variables/columnas o transforma variables existentes

summarise/summarize: genera resúmenes estadísticos de diferentes variables en el data frame, posiblemente con *strata*

%>% : el operador "pipe" es usado para conectar múltiples acciones en una única "pipeline" (tubería)