

```
library(dplyr)

rladies_global %>%
  filter(city == 'Lima')
```



R-LADIES LIMA

¡Que empiece el juego!

Primeros pasos en R y RStudio





¡Hola mundo!

Soy Fiorella Flores

Estoy en 10mo ciclo de Ing. Estadística y
me encanta programar *_*

Twitter: @FioreFloresC

Entorno de RStudio y Configuración



Interfaz que permite acceder de manera sencilla a todos los procedimientos que se realizan en R.

Actualmente es la interfaz más empleada por los usuarios.

Entorno de RStudio y Configuración



Entorno de RStudio

Consta de 4 paneles:

El **Editor**, donde se escriben los comandos.

La **Consola**, donde los comandos son ejecutados.

La **Salida**, donde se muestran los resultados gráficos.

Environment/History, donde se muestran los elementos creados durante la sesión.

Entorno de RStudio y Configuración



Entorno de RStudio

The screenshot displays the RStudio environment with four main panes:

- Editor:** Contains a script named 'Untitled1.R' with the following R code:

```
1 x<-1:10
2 y<-rnorm(10)
3 plot(x,y)
4
```
- Console:** Shows the R version and platform information, followed by the R license text and the execution of the code from the script:

```
R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

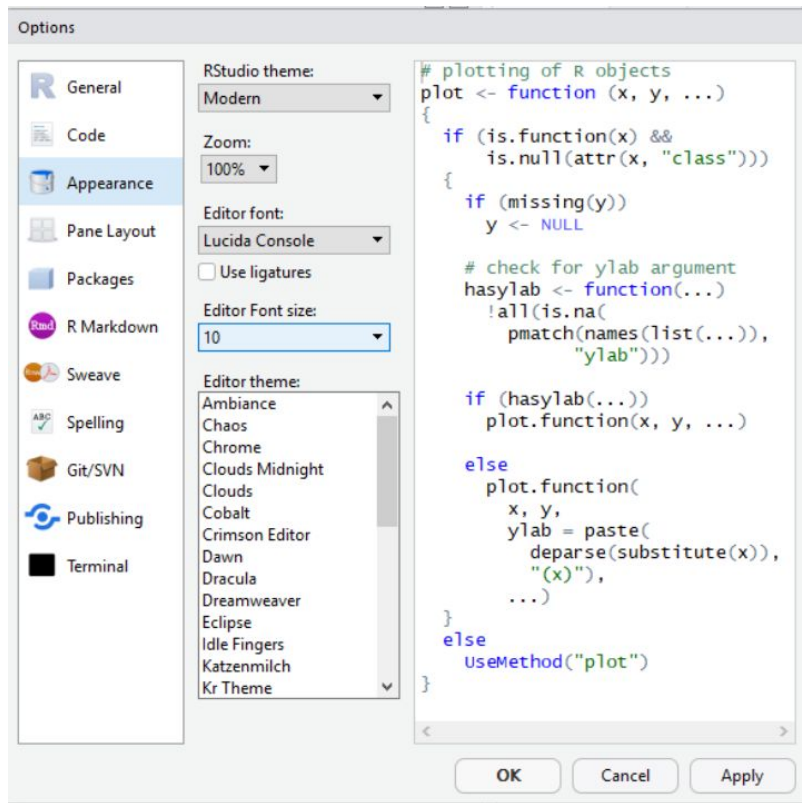
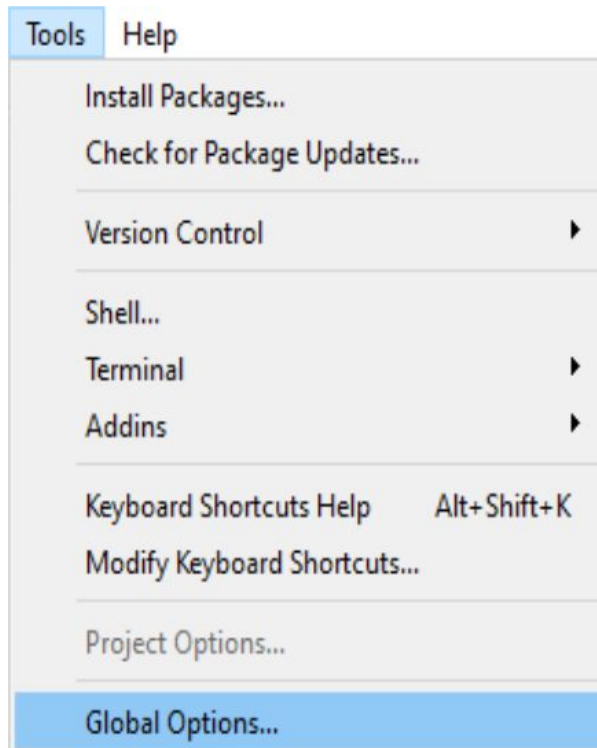
> x<-1:10
> y<-rnorm(10)
> plot(x,y)
>
```
- Environment:** Displays the current environment with variables 'x' and 'y'.

| Variable | Class | Values |
|----------|------------|---|
| x | int [1:10] | 1 2 3 4 5 6 7 8 9 10 |
| y | num [1:10] | -0.1288 -0.0539 -0.347 1.0532 -1.6641 ... |
- Plots:** Displays a scatter plot titled 'Salida' showing the relationship between 'x' and 'y'. The x-axis ranges from 1 to 10, and the y-axis ranges from -2.0 to 1.0. The plot shows 10 data points scattered around the origin.

Entorno de RStudio y Configuración



Configuración: Apariencia del Entorno



Entorno de RStudio y Configuración



Configuración: Directorio de trabajo

R siempre considera un directorio específico de trabajo. Para conocer el actual directorio de trabajo, se ejecuta la función `getwd()`.

```
Console Terminal x
~/R/ ↗
> getwd()
[1] "C:/Users/User/Documents/R"
>
```

Para cambiar de directorio, se ejecuta la función `setwd()` especificando la ruta completa.

```
> setwd("C:/Users/User/Desktop/Taller Intro al R")
> getwd()
[1] "C:/Users/User/Desktop/Taller Intro al R"
>
```

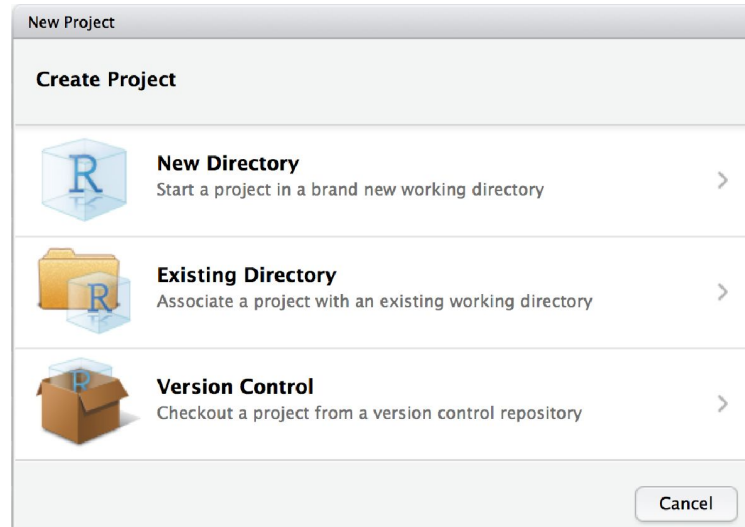
Entorno de RStudio y Configuración



Iniciar un Proyecto

Se recomienda trabajar dentro de un proyecto en RStudio, de modo que se pueda tener un propio directorio, workspace, scripts e historial en él.

Para crear un nuevo proyecto, se hace click en File – New Project. A continuación, elegir un nombre y ubicación.



Instalación y Uso de Paquetes



Instalación de Paquetes

Desde el CRAN:

Hacer click en Tools – Install Packages, o escribir en la consola: `install.packages("Nombre_del_paquete")`.

Desde GitHub:

Instalar previamente desde CRAN el paquete “devtools”.

Escribir `library(devtools)`.

Escribir `install_github("ubicación del paquete dentro de GitHub")`.

Uso de Paquetes

Escribir en la consola: `library(nombre_del_paquete)`

Escribir `?nombre_del_paquete` o `help(nombre_del_paquete)` para conocer las funciones y conjuntos de datos dentro del paquete.

Creación de Objetos



| Tipos de Datos |
|---|
| Numeric (Números reales) |
| Integer (Números enteros) |
| Complex (Números complejos) |
| Logical (Lógicos: Verdadero – TRUE / Falso – FALSE) |
| Character (caracteres, palabras) |

Creación de Objetos



Los datos en R se almacenan en objetos a través del operador de Asignación:

Objeto de
nombre "a"

a <-

Operador de
Asignación

| Dióxido de Nitrógeno [µg/m³] | | | | | | | | | | | | | | | | | | | | | | | | 13/10/2017 |
|------------------------------|-----|-----|-----|-----|----|----|----|----|-----|-----|-----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|------------|
| Estación / Hora | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| Pza. de España | 99 | 85 | 74 | 65 | 60 | 53 | 55 | 68 | 86 | 99 | 102 | 88 | 81 | 69 | 54 | 41 | 25 | 35 | 48 | 91 | 107 | 125 | 125 | 113 |
| Escuelas Aguirre | 93 | 69 | 54 | 39 | 45 | 50 | 55 | 91 | 100 | 89 | 70 | 68 | 74 | 69 | 65 | 54 | 52 | 59 | 79 | 94 | 134 | 160 | 158 | 123 |
| Ramón y Cajal | 60 | 31 | 20 | 13 | 13 | 16 | 31 | 56 | 85 | 73 | 42 | 54 | 61 | 74 | 67 | 53 | 28 | 30 | 40 | 67 | 146 | 153 | 121 | 137 |
| Arturo Soria | 42 | 22 | 16 | 16 | 16 | 20 | 25 | 53 | 69 | 68 | 47 | 39 | 46 | 52 | 40 | 34 | 25 | 27 | 36 | 63 | 110 | 166 | 164 | 138 |
| Villaverde | 147 | 133 | 115 | 102 | 94 | 90 | 90 | 95 | 129 | 131 | 103 | 81 | 52 | 45 | 32 | 19 | 17 | 22 | 35 | 83 | 141 | 142 | 146 | 134 |
| Farolillo | 95 | 87 | 80 | 73 | 62 | 61 | 58 | 58 | 71 | 78 | 87 | 63 | 68 | 52 | 38 | 18 | 17 | 19 | 29 | 63 | 111 | 129 | 124 | 106 |
| Casa de Campo | 88 | 61 | 50 | 45 | 37 | 30 | 21 | 27 | 24 | 30 | 51 | 42 | 50 | 59 | 31 | 12 | 4 | 3 | 3 | 10 | 41 | 78 | 74 | 73 |
| Barajas | 77 | 56 | 38 | 44 | 58 | 67 | 71 | 88 | 83 | 72 | 68 | 52 | 41 | 47 | 37 | 21 | 19 | 19 | 26 | 67 | 119 | 134 | 130 | 100 |
| Pza. del Carmen | 122 | 94 | 86 | 67 | 78 | 67 | 73 | 86 | 89 | 94 | 84 | 77 | 76 | 73 | 60 | 41 | 41 | 43 | 50 | 70 | 98 | 111 | 106 | 106 |
| Moratalaz | 56 | 28 | 19 | 18 | 20 | 25 | 39 | 75 | 77 | 84 | 63 | 52 | 70 | 56 | 43 | 36 | 23 | 27 | 37 | 47 | 111 | 116 | 113 | 111 |
| Cuatro Caminos | 64 | 55 | 41 | 29 | 38 | 63 | 49 | 56 | 78 | 67 | 76 | 57 | 62 | 63 | 48 | 36 | 23 | 23 | 31 | 56 | 88 | 126 | 101 | 98 |
| Barrio Pilar | 100 | 71 | 40 | 25 | 20 | 42 | 57 | 82 | 85 | 76 | 55 | 47 | 53 | 69 | 49 | 47 | 35 | 35 | 51 | 94 | 133 | 149 | 145 | 133 |
| Pte. de Vallecas | 60 | 40 | 30 | 22 | 23 | 28 | 49 | 85 | 86 | 73 | 61 | 55 | 49 | 41 | 30 | 25 | 15 | 19 | 36 | 48 | 75 | 116 | 117 | 115 |
| Méndez Alvaro | 99 | 90 | 84 | 77 | 69 | 58 | 61 | 69 | 79 | 90 | 76 | 52 | 60 | 48 | 30 | 18 | 10 | 12 | 19 | 57 | 132 | 160 | 154 | 123 |
| Castellana | 55 | 37 | 25 | 18 | 15 | 16 | 24 | 57 | 70 | 68 | 46 | 46 | 60 | 62 | 53 | 43 | 36 | 38 | 43 | 74 | 119 | 115 | 121 | 115 |
| Retiro | 93 | 51 | 35 | 27 | 20 | 20 | 22 | 46 | 59 | 65 | 62 | 57 | 63 | 51 | 33 | 21 | 10 | 12 | 18 | 22 | 52 | 75 | 88 | 92 |
| Pza. de Castilla | 61 | 31 | 24 | 19 | 17 | 19 | 33 | 55 | 77 | 61 | 32 | 30 | 41 | 55 | 45 | 41 | 40 | 45 | 61 | 87 | 124 | 118 | 109 | 87 |
| Ens. de Vallecas | 69 | 50 | 49 | 44 | 43 | 58 | 79 | 94 | 84 | 70 | 58 | 52 | 50 | 44 | 21 | 17 | 15 | 14 | 19 | 47 | 128 | 161 | 156 | 142 |
| Urb. Embajada | 74 | 53 | 46 | 43 | 66 | 76 | 81 | 97 | 98 | 88 | 78 | 46 | 49 | 57 | - | 25 | 20 | 21 | 33 | 71 | 100 | 115 | 138 | 123 |
| Fdez. Ladreda | 109 | 101 | 94 | 84 | 73 | 73 | 72 | 92 | 103 | 108 | 97 | 91 | 81 | 66 | 56 | 36 | 31 | 36 | 73 | 147 | 229 | 224 | 126 | 120 |
| Sanchinarro | 64 | 24 | 19 | 11 | 9 | 10 | 20 | 40 | 69 | 58 | 39 | 26 | 27 | 36 | 35 | 29 | 16 | 12 | 23 | 52 | 115 | 117 | 88 | 109 |
| El Pardo | 22 | 20 | 16 | 14 | 12 | 11 | 19 | 35 | 27 | 25 | 19 | 16 | 16 | 26 | 33 | 18 | 5 | 4 | 9 | 48 | 43 | 36 | 31 | 20 |
| Juan Carlos I | 24 | 13 | 9 | 9 | 8 | 11 | 26 | 40 | 42 | 27 | 24 | 22 | 24 | - | 26 | 19 | 12 | 11 | 14 | 34 | 82 | 110 | 115 | 108 |
| Tres Olivos | 94 | 47 | 34 | 18 | 14 | 15 | 24 | 61 | 83 | 73 | 55 | 37 | 35 | 46 | 40 | 33 | 17 | 17 | 20 | 45 | 89 | 97 | 95 | 91 |

Creación de Objetos



Algunas indicaciones para nombrar Objetos

- ❖ Usar letras del alfabeto inglés, números, punto y guión bajo.
- ❖ No colocar números y puntos al comienzo del nombre.
- ❖ Tener en cuenta que “X” y “x” son diferentes nombres.
- ❖ No asignar nombres que ya identifican a funciones conocidas, palabras reservadas (especialmente T, F, NA, NaN, Inf y NULL) y objetos predefinidos como pi, LETTERS y letters. Para ver todas las palabras reservadas, escribir `?reserved` en la consola.

Creación de Objetos



Tipos de Objetos

Los objetos atómicos son aquellos cuyas componentes son todas de tipo numeric, character, logical o complex, como los **vectores** y las **matrices**.

Vectores

Funciones para construir vectores:

| | | |
|--------------------|-----------------|------------------------------|
| Función Concatenar | c(...) | Combina valores. |
| Función seq | seq(from,to,by) | Genera secuencias numéricas. |
| Función : | from:to | Genera secuencias numéricas. |
| Función rep | rep(x,times) | Repite valores. |

Creación de Objetos



Matrices

Se usa la función `matrix()` para crear una matriz.

Argumentos:

| | |
|-------|--|
| data | Un vector de datos. |
| nrow | Número de filas. |
| ncol | Número de columnas. |
| byrow | Si es TRUE, la matriz es llenada por filas. Si es FALSE, la matriz es llenada por columnas. |

Creación de Objetos



Data Frames

Pueden estar formados de diferentes tipos de objetos: vectores, matrices u otros data frames.
Se define con la función `data.frame()`

```
> df1<-data.frame(nombres=c("Clase1","Clase2","Clase3"),numeros=c(3,4,6),  
+                  numeros1=c(3,4,1))
```

```
> df1
```

| | nombres | numeros | numeros1 |
|---|---------|---------|----------|
| 1 | Clase1 | 3 | 3 |
| 2 | Clase2 | 4 | 4 |
| 3 | Clase3 | 6 | 1 |

Creación de Objetos



Listas

Puede contar con elementos de distintas clases y de diferentes dimensiones. Se usa la función *list()*.

```
> lista
[[1]]
[1] 1 2 3 4 5 6 7 8 9 10

[[2]]
      [,1] [,2] [,3]
[1,]    2    6   10
[2,]    3    7   11
[3,]    4    8   12
[4,]    5    9   13

[[3]]
      Petal.Length Petal.width Species
10             1.5         0.1  setosa
11             1.5         0.2  setosa
12             1.6         0.2  setosa
13             1.4         0.1  setosa
14             1.1         0.1  setosa
15             1.2         0.2  setosa
16             1.5         0.4  setosa

$mes
[1] "marzo"
```


Creación de Objetos



Inspección de Objetos

| Función | Utilidad |
|-----------------------|--|
| <code>class()</code> | Obtiene el tipo o modo de almacenamiento de un objeto. |
| <code>length()</code> | Obtiene la longitud de un objeto. |
| <code>dim()</code> | Obtiene la dimensión de un objeto. |
| <code>head()</code> | Muestra las 6 primeras filas o elementos de un objeto. |
| <code>str()</code> | Muestra la estructura de un objeto. |
| <code>ls()</code> | Muestra los nombres de todos los objetos creados. |
| <code>rm()</code> | Borra el objeto. |

Importación y exportación de datos



Función write.table()

Guarda un objeto con extensión .rdata.

El objeto a guardar debe ser preferiblemente una matriz o un data frame. De no ser así, se intentará forzar al objeto a ser un data frame.

```
> write.table(df1, file = "df1taller.rdata")
```

Función read.table()

Lee el objeto guardado con extensión .rdata y crea un data frame.

También puede leer archivos con otro tipo de extensiones, como .txt y .csv.

```
> df1taller<-read.table("df1taller.rdata")
> df1taller
  nombres numeros1 numeros2
1  Clase1         3         3
2  Clase2         4         4
3  Clase3         6         1
```

Importación y exportación de datos



Funciones read.csv() y read.csv2()

Leen hojas de cálculo con extensión .csv (Comma Separated Values)

La función read.csv() tiene como símbolo de separación una coma (,), mientras la función read.csv2() usa un punto y coma (;)

```
> fiscalizacion <- read.csv("Fiscalizacion.csv", sep=",", header = TRUE)
```

Función read.xls()

Lee archivos Excel (extensión .xls o .xlsx) y los almacena en un data frame en R.
Requiere la instalación del paquete “gdata”.

```
> install.packages("gdata")
```

```
> library(gdata)
```

```
> fisc<-read.xls("Fiscalizacion.xlsx", sheet = 1)
```

Operaciones matemáticas básicas y estadística descriptiva



Operadores en R

| Aritméticos | | Operadores Comparativos | | Lógicos | |
|-------------|---------------------|-------------------------|-------------------|------------|-------------|
| + | adición | < | menor que | ! x | NO lógico |
| - | substracción | > | mayor que | x & y | Y lógico |
| * | multiplicación | <= | menor o igual que | x && y | id. |
| / | división | >= | mayor o igual que | x y | O lógico |
| ^ | potencia | == | igual | x y | id. |
| % % | módulo | != | diferente de | xor (x, y) | O exclusivo |
| % / % | división de enteros | | | | |

Operaciones matemáticas básicas y estadística descriptiva



Estadística Descriptiva

| Función | Utilidad |
|---------|---|
| mean() | Calcula la media aritmética de un vector. |
| var() | Calcula la variancia de un vector. |
| sd() | Calcula la desviación estándar de un vector. |
| min() | Calcula el valor mínimo de un conjunto de datos. |
| max() | Calcula el valor máximo de un conjunto de datos. |
| range() | Calcula el valor mínimo y máximo de un conjunto de datos. |

Operaciones matemáticas básicas y estadística descriptiva



Estadística Descriptiva

| Función | Utilidad |
|-------------------------|---|
| <code>cor()</code> | Calcula la correlación entre dos vectores x e y . |
| <code>median()</code> | Calcula la mediana de un conjunto de datos. |
| <code>quantile()</code> | Obtiene diversos percentiles. |
| <code>summary()</code> | Obtiene estadísticas de resumen. |

Gráficos estadísticos



Histograma

Principales argumentos:

| | |
|--------|---|
| x | Vector de valores. |
| breaks | Indica el número aproximado de clases, también puede ser un vector con los puntos límites. |
| freq | Si es T se presentan las frecuencias absolutas, si es F se presentan las frecuencias relativas. |
| xlab | Etiqueta para el eje x. |
| ylab | Etiqueta para el eje y. |
| xlim | Límites para el eje x. |
| ylim | Límites para el eje y. |

Gráfica de cajas o Box Plot

Principales argumentos:

| | |
|------------|--|
| x | Vector de datos que van a ser graficados. |
| range | Determina la extensión de los bigotes de la gráfica de caja. |
| horizontal | Valor lógico que indica si la caja debería ser horizontal, por defecto es FALSE. |
| col | Indica el color de la caja. |
| boxwex | Un factor de escala aplicado a todas las cajas. |
| main | Título principal. |
| xlab | Etiqueta para el eje x. |
| ylab | Etiqueta para el eje y. |

Gráficos estadísticos



Gráfica de barras

Principales argumentos:

| | |
|---------|---|
| height | Un vector o matriz de valores que describen las barras. |
| col | Un vector de colores para las barras o los componentes de barras. |
| main | Título principal. |
| xlab | Etiqueta para el eje x. |
| ylab | Etiqueta para el eje y |
| xlim | Límites para el eje x. |
| ylim | Límites para el eje y |
| beside | Un valor lógico. Si es T, las columnas serán barras yuxtapuestas. Por defecto es F (barras apiladas). |
| density | Un vector que da la densidad de líneas de sombreado. |

Gráfica circular

Principales argumentos:

| | |
|---------|--|
| x | Un vector de cantidades numéricas no negativas. |
| labels | Una o más expresiones de cadenas de caracteres que dan los nombres de las áreas. |
| col | Un vector de colores a ser usados para el llenado. |
| density | Un vector que da la densidad de líneas de sombreado. |

Ejercicio Práctico



- ☐ Crear una secuencia del 1 al 50 de 2 en 2 y almacenarlo en un objeto.
- ☐ Hallar la cantidad de datos (longitud), promedio y mediana del objeto.
- ☐ Extraer el 4to valor de la secuencia.

Se tiene el dataset "cars" que contiene 50 observaciones que registran la velocidad de los carros y las distancias que tomaron en detenerse.

Sus 2 variables se encuentran almacenadas en:

```
velocidad<-cars$speed  
distancia<-cars$dist
```

- ☐ Describir la estructura del dataset y hallar su correlación.
- ☐ Hallar el promedio, desviación estándar y variancia de cada variable.
- ☐ Graficar el histograma de "velocidad" y boxplot de "distancia".