

Sesión 1:

El Lenguaje R:
Operaciones y
Gráficas Básicas



Objetivos

- Realizar una introducción al software R y sus características básicas.
- Presentar algunas técnicas y herramientas de manejo de datos y análisis estadístico presentes en R.
- Visualizar información estadística mediante R.

Palabras claves: R, R-Studio, programación, variable, entorno, código, dataframe.

Introducción

Los procesos de inteligencia de negocios exigen el manejo de muchos datos: los datos deben ser limpiados, organizados y filtrados antes de poder realizarse algún análisis estadístico o usarlos en procesos de aprendizaje de máquina.

Considerando la gran cantidad de datos que podrían utilizarse en un solo proceso, es conveniente que el manejo de los datos pueda realizarse de forma algorítmica, para garantizar velocidad de procesamiento y seguridad de la información.

Para tal fin se usa software de programación, entre los cuales se encuentra R, un proyecto de libre acceso diseñado para el análisis de datos estadísticos.

1. R y R-Studio

Como se menciona en su página web, "R es un lenguaje y entorno para la computación estadística y gráfica" (R: The R Project for Statistical Computing, s. f.). R no solo consta del lenguaje de programación, sino que además tiene un sistema de almacenamiento, herramientas de cálculo y análisis de datos, librerías gráficas para visualizaciones, entre otras.

Se considera un entorno al incluir todas las herramientas necesarias para su implementación y manejo. Tiene además un gran repositorio de paquetes diseñados para necesidades específicas y una comunidad de usuarios activa y colaboradora.

Es posible programar en R desde una terminal, pero por comodidad se usa un entorno gráfico (R Studio) que es más intuitivo y permite manejar simultáneamente códigos, archivos, bases de datos y visualización.

a. Instalación:

El paquete básico de R se encuentra disponible en la página CRAN, Comprehensive R Archive Network (CRAN - Mirrors, s. f.), donde se encuentra el repositorio de descargas.

Para empezar, se elige un repositorio y a continuación se elige la versión deseada del programa como lo muestra la Ilustración 1.

Chile	https://cran.dcc.uchile.cl/	Departamento de Ciencias de la Computación, Universidad de Chile
China	https://mirrors.tuna.tsinghua.edu.cn/CRAN/ https://mirrors.bfsu.edu.cn/CRAN/ https://mirrors.ustc.edu.cn/CRAN/ https://mirror.hk.koddos.net/CRAN/ https://mirrors.e-education.cn/CRAN/ https://mirror.lzu.edu.cn/CRAN/ https://mirrors.nju.edu.cn/CRAN/ https://mirrors.tongji.edu.cn/CRAN/ https://mirrors.sjtu.edu.cn/cran/ https://mirrors.sustech.edu.cn/CRAN/	TUNA Team, Tsinghua University Beijing Foreign Studies University University of Science and Technology of China KoDDOs in Hong Kong Elite Education Lanzhou University Open Source Society eScience Center, Nanjing University Tongji University Shanghai Jiao Tong University Southern University of Science and Technology (SUSTech)
Colombia	https://www.icesi.edu.co/CRAN/	Icesi University
Costa Rica	https://mirror.uned.ac.cr/cran/	Distance State University (UNED)
Czech Republic	https://mirrors.nic.cz/R/	CZ.NIC, Prague
Denmark	https://mirrors.dotsrc.org/cran/	Aalborg University

Ilustración 1.R. CRAN Mirrors. Recuperado de <https://cran.r-project.org/mirrors.html>



The Comprehensive R Archive Network

Download and Install R
Precompiled binary distributions of the base system and contributed packages. Windows and Mac users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for Mac OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms
Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

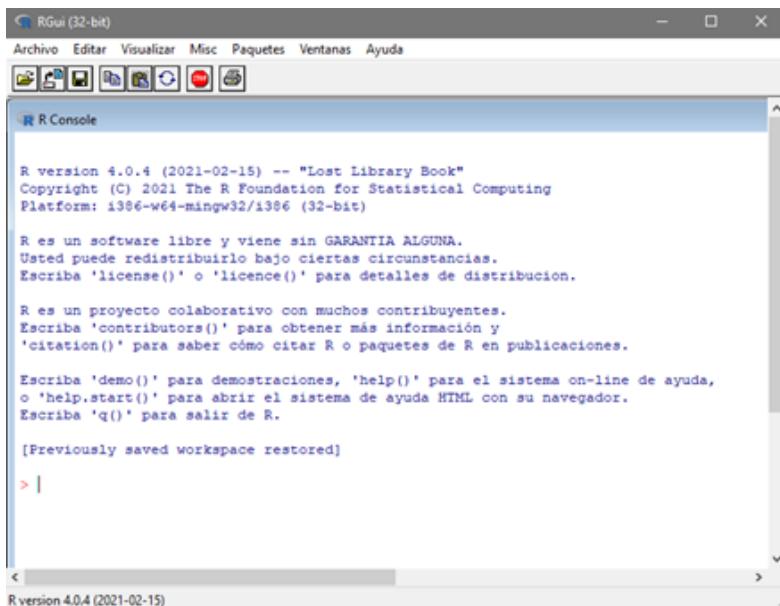
- The latest release (2021-03-31, Shake and Throw) [R-4.0.5.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extensive [packages](#)

Questions About R
• If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

What are R and CRAN?

Ilustración 2. R. The Comprehensive R Archive Network. Recuperado de <https://cran.r-project.org/>

En Windows (<https://www.icesi.edu.co/CRAN/bin/windows/base/>) como se muestra en la Ilustración 2, se obtiene un instalador que trae los paquetes básicos de R y una interfaz sencilla tipo terminal representado por la ilustración 3.



R Gui (32-bit)

Archivo Editar Visualizar Misc Paquetes Ventanas Ayuda

R Console

```
R version 4.0.4 (2021-02-15) -- "Lost Library Book"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-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.

[Previously saved workspace restored]

> |
```

version 4.0.4 (2021-02-15)

Ilustración 3. Consola de R

Se puede empezar a programar desde esta terminal, pero para efectos prácticos una interfaz de usuario más amigable es recomendable(RStudio) como lo muestra la Ilustración 4.

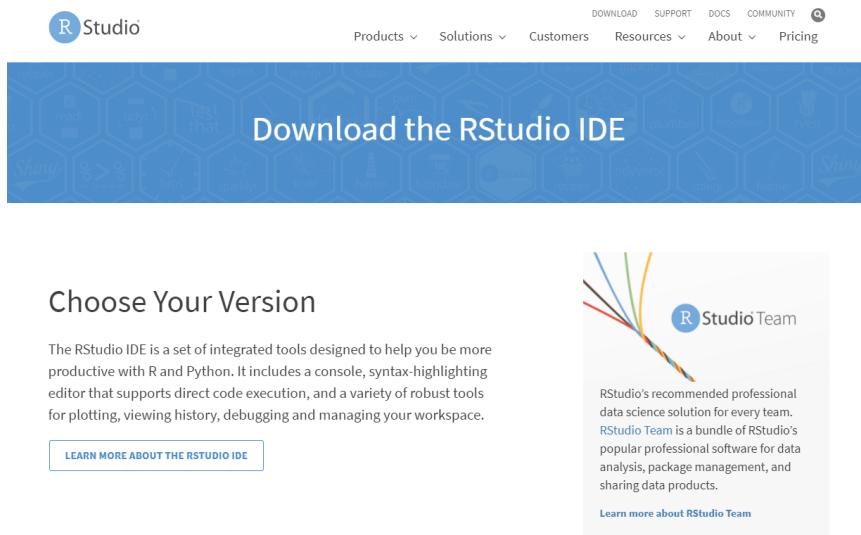


Ilustración 4. RStudio. Download the RStudio IDE. Recuperado de <https://www.rstudio.com/products/rstudio/-/download/>

R-Studio(Download the RStudio IDE, s. f.) es un conjunto de herramientas diseñadas para integrarse con R y Python. Además de la consola, integra el manejo de las bases de datos y visualizaciones junto a un gestor de librerías.

Trabaja sobre la base de R, así que su instalación debe ser posterior a la instalación del paquete básico R. Una vez descargado y ejecutado el instalador, RStudio se sincroniza con la versión ya instalada de R.

La interfaz de RStudio se puede organizar en paneles a través de Menú View-Panes-Pane Layout como lo muestra la ilustración 5. Generalmente se usan cuatro paneles: Source (Fuente), donde se escribe el código a ejecutar; Console (Consola), donde se ejecuta el código y se observan resultados; Environment (Entorno), donde se pueden ver los archivos actualmente en uso; Plots (Graficas) donde aparecen las

gráficas construidas por el código una vez ejecutado en consola. La interfaz es personalizable con más o menos pestañas en cada panel de acuerdo a lo que necesitemos. Valga la pena anotar que, debido a su arquitectura, R necesita una buena máquina para funcionar correctamente. Trabajando con muchos datos puede experimentar demoras.

Una alternativa es RCloud, una implementación online de R que trabaja en nube. Mientras los datos que se usen no sean confidenciales, es una buena alternativa para máquinas con pocos recursos.

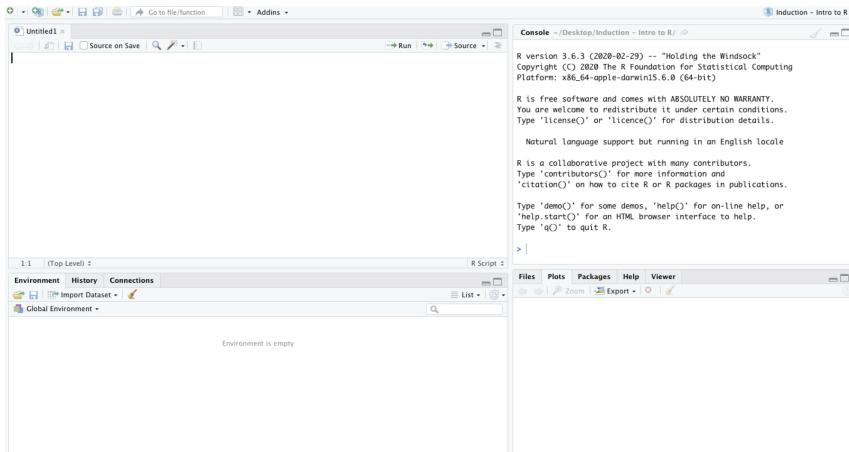


Ilustración 5. Interfaz de RStudio

2. Datos en R y Operaciones:

Como se menciona en Santana & Farfán (s. f.), R es un lenguaje de programación orientado a objetos, lo que lo hace bastante eficiente para las tareas de análisis y procesamiento.

Dichos objetos pueden combinarse para generar objetos nuevos, a partir de la base de clases atómicas: character, numeric, integer, complex, logical. Las funcionalidades básicas de R combinadas con las prestaciones de RStudio, los convierten en un sistema muy robusto para análisis de datos.

Como un ejemplo de la facilidad de RStudio para trabajar, se pueden mencionar los proyectos: un único archivo que enlaza scripts, variables y bases de datos para que al iniciar una sesión de trabajo se puedan conectar todas las piezas necesarias.

Actividad 1: Cree un proyecto para la sesión.

En la interfaz de RStudio se pueden ejecutar líneas de código en la consola, pero es más común y efectivo mantenerlas todas dentro de un script.

Dentro de los scripts, las líneas marcadas con # al comienzo se toman como comentarios y se ignoran durante la ejecución. Una línea específica del código puede ejecutarse en la consola (garantizando que todo a lo que haga esa línea ya haya sido ejecutado), pero RStudio permite ejecutar una línea específica del script presionando Ctrl+Enter con el cursor sobre una línea en particular.

Actividad 2: Ejecute la siguiente línea y verifique que el resultado de la operación aparece en la consola.

```
10+3
```

```
## [1] 13
```

R se ocupa de todos los cálculos básicos de una calculadora, como se ilustra con las siguientes líneas.

```
log(4)
```

```
## [1] 1.386294
```

```
sin(pi)
```

```
## [1] 1.224606e-16
```

Nótese que en esta última línea está realizando una aproximación del resultado, que es cero.

Aparte de las operaciones básicas, se pueden construir algunas estructuras más interesantes. Recordando las clases básicas de R.

```
class(4)  
  
## [1] "numeric"
```

Aunque 4 es un numero entero, R lo reconoce en clase numérica. Para pasarlo como entero se puede agregar una L al final o usar la función as.integer.

```
class(as.integer(4))  
  
## [1] "integer"
```

3. Manejo de datos

Se pueden componer funciones en R de manera anidada. Si queremos guardar algún resultado o valor en una variable, la asignamos con los operadores -> o <-.

```
x<- 4+7  
x  
## [1] 11  
  
class(x)  
  
## [1] "numeric"
```

Guardar este valor en la variable x, hace que R la reconozca como un vector: un conjunto de elementos de la misma naturaleza.

El [1] que sale en la consola hace referencia al primer elemento del vector. Para crear vectores de mayor tamaño se puede usar la función c()

```
x<- c(4,3,-2)  
x  
  
## [1] 4 3 -2
```

```
class(x)  
  
## [1] "numeric"
```

¿Qué pasa si se crea un vector con objetos de distinto tipo?

```
x<-c(3,'holo')  
class(x)  
  
## [1] "character"
```

Por un proceso de coerción, R trata de dejar todos los objetos de un vector de la misma clase. Cuando los vectores son de tipo numérico, entero o complejo, se pueden realizar operaciones componente a componente

```
x<-c(4,5,2)  
x*2  
  
## [1] 8 10  4  
  
y<-c(-3,2,1)  
x+y  
  
## [1] 1 7  3  
  
x*y  
  
## [1] -12 10   2
```

En esta última, la operación que se realizó fue multiplicar componente a componente. Si se quiere la multiplicación de matrices (producto punto) debe usarse %*%

```
x%*%y  
  
##      [,1]  
## [1,]    0
```

Si los vectores que se están usando para la operación no son de tamaños iguales, R reutiliza valores para completar las operaciones

```
x<-c(4,2,5,8,9)
```

```
x+y
```

```
## Warning in x + y: longitud de objeto mayor no es
múltiplo de la longitud de uno
## menor
## [1] 1 4 6 5 11
```

Sobre un objeto se pueden aplicar pruebas lógicas para determinar condiciones

```
z<-x>=5
```

```
z
```

```
## [1] FALSE FALSE TRUE TRUE TRUE
class(z)
## [1] "logical"
```

En un vector es posible llamar elementos específicos usando el índice y modificarlos. Además, se puede calcular la longitud del vector con length()

```
x[3]<--1
```

```
x
```

```
## [1] 4 2 -1 8 9
```

```
length(x)
```

```
## [1] 5
```

El entorno de ejecución guarda los valores actuales de las variables: Si alguna variable se actualizara y se ejecutara una línea anterior que la referencia, la ejecución se haría con el ultimo valor.

Aunque los objetos actuales están listados en el entorno, se pueden ver con el comando ls().

```
ls()
```

```
## [1] "x" "y" "z"
```

El comando rm() permite borrar objetos. También es posible construir una matriz, puede hacerse con el comando matrix, especificando el vector que se quiere convertir en una matriz.

```
A<-matrix(data=c(3,2,1,5,6,1,-2,4), nrow=2, ncol=4)
```

```
A
```

```
##      [,1] [,2] [,3] [,4]
## [1,]     3     1     6    -2
## [2,]     2     5     1     4
```

En esta última los valores se organizaron por columnas. Si se quieren organizar por filas se agrega un parámetro

```
A<-matrix(data=c(3,2,1,5,6,1,-2,4), nrow=2, ncol=4,
byrow=T)
```

```
A
```

```
##      [,1] [,2] [,3] [,4]
## [1,]     3     2     1     5
## [2,]     6     1    -2     4
```

También se pueden usar vectores separados y unirlos por filas o columnas

```
A<-rbind(c(3,2,1),c(5,7,-1))
```

```
A
```

```
##      [,1] [,2] [,3]
## [1,]     3     2     1
## [2,]     5     7    -1
```

```
class(A)
## [1] "matrix" "array"
```

Actividad 3: Construya una matriz 4x4 con los números naturales del 1 al 16, organizada por filas.

La actividad anterior es bastante tediosa con las herramientas hasta ahora descritas, pero se puede crear una secuencia de valores de manera rápida para simplificar el problema.

La expresión a:b crea una secuencia que empieza en a y se incrementa en una unidad hasta alcanzar o sobrepasar a b

```
s<-3:8.5  
s  
  
## [1] 3 4 5 6 7 8  
  
class(s)  
  
## [1] "integer"
```

Si se quiere cambiar el tamaño de paso, se puede usar esta otra estructura

```
s<- seq(from=3, to=8.5, by=0.5)  
s  
  
## [1] 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5  
  
class(s)  
  
## [1] "numeric"
```

La matriz de anterior entonces se puede construir como

```
matrix(1:16, nrow=4, ncol=4, byrow=T)  
  
##      [,1] [,2] [,3] [,4]  
## [1,]     1     2     3     4
```

```
## [2,]    5    6    7    8
## [3,]    9   10   11   12
## [4,]   13   14   15   16
```

También es conveniente mencionar las listas. A diferencia de los vectores estas pueden contener objetos de diversas clases.

```
l<-list(3,5,5.3,'Hola')
```

```
l
```

```
## [[1]]
## [1] 3
##
## [[2]]
## [1] 5
##
## [[3]]
## [1] 5.3
##
## [[4]]
## [1] "Hola"
```

```
class(l)
```

```
## [1] "list"
```

```
l[[4]]
```

```
## [1] "Hola"
```

```
k<-list(c(2,3),c(4,5,2))
```

```
k
```

```
## [[1]]
## [1] 2 3
```

```
##  
## [[2]]  
## [1] 4 5 2
```

En este último se tiene una lista con elementos de distinto tamaño. Es de intereses una estructura similar a una lista, pero con la condición que los elementos son todos del mismo tamaño: el dataframe.

```
d<-data.frame(c(2,3,4),matrix(-9:-1,nrow=3,ncol=3))  
d  
  
##   c.2..3..4. X1 X2 X3  
## 1      2 -9 -6 -3  
## 2      3 -8 -5 -2  
## 3      4 -7 -4 -1  
class(d)  
  
## [1] "data.frame"  
  
d[1]  
  
##   c.2..3..4.  
## 1      2  
## 2      3  
## 3      4  
  
d$X1  
  
## [1] -9 -8 -7
```

Un dataframe tiene una estructura similar a una matriz, y tiene la ventaja que se pueden sacar partes (frames) del arreglo, filas o columnas, para trabajar con ellos. Generalmente aparecen al momento de cargar bases de datos.

Actividad 4: Descargue la base de datos presente en <https://www.datos.gov.co/dataset/Caracterizaci-n-Madres-y-Padres-Comunitarios-ICBF/ixwb-pqgb> y agréguela al entorno de trabajo usando el botón Import Dataset del panel entorno. Una vez cargada, ejecute la siguiente línea

```
class(Caracterizaci_n_Madres_y_Padres_Comunitarios_IC  
BF)  
  
## [1] "data.frame"
```

Al leer una base de datos R la convierte en un dataframe. Nótese que también genera una pestaña de visualización de los datos. Los bloques básicos del dataframe son las columnas, que tienen un identificador o nombre.

Si la base tenía un encabezado, hereda el nombre. De otro modo, R lo nombra de manera automática.

```
colnames(Caracterizaci_n_Madres_y_Padres_Comunitarios  
_ICBF)  
## [1] "DEPARTAMENTOUDS"  
## [2] "MUNICIPIOUDS"  
## [3] "CENTROZONALUDS"  
## [4] "ZONAUBICACIONUDS"  
## [5] "UNIDADSERVICIO"  
## [6] "SERVICIOUDS"  
## [7] "SEXOTALENTOHUMANO"  
## [8] "EDADENAÑ.OS"  
## [9] "GRUPOETNICOTALENTOHUMANO"  
## [10] "CANTIDAD.MADRES.Y.O.PADRES.COMUNITARIOS"
```

Se puede llamar a una columna específica usando el nombre o índice y el operador `[[]]`. Con el nombre también puede usarse \$

```
Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF  
[[ 'SERVICIOUDS' ]]
```

```
Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF$EDA  
DENAÑ.OS
```

Para llamar las filas se puede hacer por nombres o índices

```
Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF[1, ]
```

```
## DEPARTAMENTOUDS MUNICIPIOUDS CENTROZONALUDS
ZONAUBICACIONUDS
## 1 ANTIOQUIA ABEJORRAL CZ ORIENTE
CABECERA MUNICIPAL
## UNIDADSERVICIO SERVICIOUDS
SEXOTALENTOHUMANO
## 1 HOGAR FAMI MADRES CON FUTURO HCB FAMI - FAMILIAR
FEMENINO
## EDADENAÃ.O GRUPOETNICOTALENTOHUMANO
## 1 49 NO SE AUTORECONOCE EN NINGUNO DE LOS
ANTERIORES
## CANTIDAD.MADRES.Y.O.PADRES.COMUNITARIOS
## 1 1
```

También pueden escoger algunas columnas o filas por bloque

```
Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF[1:4
, ]
```

```
## DEPARTAMENTOUDS MUNICIPIOUDS CENTROZONALUDS
ZONAUBICACIONUDS
## 1 ANTIOQUIA ABEJORRAL CZ ORIENTE
CABECERA MUNICIPAL
## 2 ANTIOQUIA ABEJORRAL CZ ORIENTE
CABECERA MUNICIPAL
## 3 ANTIOQUIA ABRIAQUI CZ OCCIDENTE MEDIO
CABECERA MUNICIPAL
## 4 ANTIOQUIA ABRIAQUI CZ OCCIDENTE MEDIO
CABECERA MUNICIPAL
## UNIDADSERVICIO SERVICIOUDS SEXOTALENTOHUMANO
## 1 HOGAR FAMI MADRES CON FUTURO HCB FAMI -
```

```

FAMILIAR FEMENINO
## 2 HOGAR FAMI MADRES HACIA EL FUTURO HCB FAMI -
FAMILIAR FEMENINO
## 3 ENSUEÑOS Y TERNURAS HCB FAMI -
FAMILIAR FEMENINO
## 4 LAS ABEJITAS HCB -
COMUNITARIO FEMENINO
## EDADENAÑOS GRUPOETNICOTALENTOHUMANO
## 1 49 NO SE AUTORECONOCE EN NINGUNO DE LOS
ANTERIORES
## 2 33 NO SE AUTORECONOCE EN NINGUNO DE LOS
ANTERIORES
## 3 41 NO SE AUTORECONOCE EN NINGUNO DE LOS
ANTERIORES
## 4 43 NO SE AUTORECONOCE EN NINGUNO DE LOS
ANTERIORES
## CANTIDAD.MADRES.Y.O.PADRES.COMUNITARIOS
## 1 1
## 2 1
## 3 1
## 4 1

```

Si queremos una previsualización del dataframe, podemos usar head().

```
head(Caracterizaci_n_Madres_y_Padres_Comunitarios_ICB
F)
```

```

## DEPARTAMENTOUDS MUNICIPIOUDS CENTROZONALUDS
ZONAUBICACIONUDS
## 1 ANTIOQUIA ABEJORRAL CZ ORIENTE
CABECERA MUNICIPAL
## 2 ANTIOQUIA ABEJORRAL CZ ORIENTE
CABECERA MUNICIPAL
## 3 ANTIOQUIA ABRIAQUI CZ OCCIDENTE MEDIO

```

CABECERA MUNICIPAL
4 ANTIOQUIA ABRIAQUI CZ OCCIDENTE MEDIO
CABECERA MUNICIPAL
5 ANTIOQUIA ABRIAQUI CZ OCCIDENTE MEDIO
CABECERA MUNICIPAL
6 ANTIOQUIA AMAGA CZ ABURRA SUR
CABECERA MUNICIPAL
UNIDADSERVICIO SERVICIOUDS SEXO TALENTO HUMANO
1 HOGAR FAMI MADRES CON FUTURO HCB FAMI -
FAMILIAR FEMENINO
2 HOGAR FAMI MADRES HACIA EL FUTURO HCB FAMI -
FAMILIAR FEMENINO
3 ENSUEÑOS Y TERNURAS HCB FAMI -
FAMILIAR FEMENINO
4 LAS ABEJITAS HCB - COMUNITARIO FEMENINO
5 LOS PICARITOS HCB - COMUNITARIO FEMENINO
6 CARIÑO 2 HCB - COMUNITARIO FEMENINO
EDADENAÑOS GRUPO ETNICO TALENTO HUMANO
1 49 NO SE AUTORECONOCE EN NINGUNO DE LOS ANTERIORES
2 33 NO SE AUTORECONOCE EN NINGUNO DE LOS ANTERIORES
3 41 NO SE AUTORECONOCE EN NINGUNO DE LOS ANTERIORES
4 43 NO SE AUTORECONOCE EN NINGUNO DE LOS ANTERIORES
5 38 NO SE AUTORECONOCE EN NINGUNO DE LOS ANTERIORES
6 50 NO SE AUTORECONOCE EN NINGUNO DE LOS ANTERIORES
CANTIDAD.MADRES.Y.O.PADRES.COMUNITARIOS
1 1
2 1
3 1

## 4	1
## 5	1
## 6	1

4. Generación de análisis estadísticos

También podemos usar una función descriptora para sacar algo de información de las variables

```
summary(Caracterizaci_n_Madres_y_Padres_Comunitarios_
ICBF)

##  DEPARTAMENTOUDS      MUNICIPIOUDS
CENTROZONALUDS      ZONAUBICACIONUDS
##  Length:47344      Length:47344
Length:47344      Length:47344
##  Class :character  Class :character  Class
:character  Class :character
##  Mode   :character  Mode   :character  Mode
:character  Mode   :character
##
##
##
##  UNIDADSERVICIO      SERVICIOUDS
SEXOTALENTOHUMANO    EDADENAÃ.OS
##  Length:47344      Length:47344
Length:47344      Min.   : 11.00
##  Class :character  Class :character  Class
:character  1st Qu.: 40.00
##  Mode   :character  Mode   :character  Mode
:character  Median  : 49.00
##                                     Mean   : 48.71
##                                     3rd Qu.: 57.00
##                                     Max.   :120.00
##  GRUPOETNICOTALENTOHUMANO
```

```
CANTIDAD.MADRES.Y.O.PADRES.COMUNITARIOS
##  Length:47344                  Min.   :1.000
##  Class :character              1st Qu.:1.000
##  Mode   :character             Median :1.000
##                                         Mean    :1.005
##                                         3rd Qu.:1.000
##                                         Max.    :3.000
```

Actividad 5: Desde la página de datos abiertos de su ciudad, descargue una base que contenga variables cualitativas y cuantitativas. Agréguela al entorno y liste sus columnas. También aplique la función summary para verificar el tipo de dato en las columnas.

La función summary dice que muchas de las columnas con las que estamos trabajando son tipo carácter. Si se selecciona una de estas y se aplica class,

```
class(Caracterizaci_n_Madres_y_Padres_Comunitarios_IC
BF$DEPARTAMENTOUDS)

## [1] "character"
```

Sobre un dataframe se pueden agregar nuevas columnas, realizando operaciones entre las columnas disponibles dependiendo de los tipos de valores en cada columna.

```
Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF$Loc
acion<-paste(Caracterizaci_n_Madres_y_Padres_Comunita
rios_ICBF$DEPARTAMENTOUDS,Caracterizaci_n_Madres_y_Pa
dres_Comunitarios_ICBF$MUNICIPIOUDS,sep=' - ')
Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF$Loc
acion[1:5]

## [1] "ANTIOQUIA-ABEJORRAL" "ANTIOQUIA-ABEJORRAL"
## [3] "ANTIOQUIA-ABRIAQUI"
## [4] "ANTIOQUIA-ABRIAQUI"  "ANTIOQUIA-ABRIAQUI"
```

Actividad 6: Con la base de datos agregada en la actividad anterior, cree una nueva columna aplicando una operación sobre una de las columnas existentes.

5. Medidas estadísticas y representación gráfica.

Sobre las columnas del dataframe pueden aplicarse operaciones estadísticas para extraer información general de las variables.

Por defecto, R trae implementadas las medidas estadísticas más generales: media, mediana, cuartiles, varianza y desviación estándar.

```
mean(Caracterizaci_n_Madres_y_Padres_Comunitarios_ICB  
F$EDADENAÃ.OS)  
  
## [1] 48.71409  
  
median(Caracterizaci_n_Madres_y_Padres_Comunitarios_I  
CBF$EDADENAÃ.OS)  
  
## [1] 49  
  
var(Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF  
$EDADENAÃ.OS)  
  
## [1] 131.7136  
  
sd(Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF$  
EDADENAÃ.OS)  
  
## [1] 11.47665
```

También es posible calcular otras medidas como la correlación. La correlación es una métrica estadística que indica cuánto se relacionan dos variables.

Dado que el dataframe actual solo tiene una variable cuantitativa, se definen dos vectores para el cálculo:

```
a<-c(4,5,6,7,8)
```

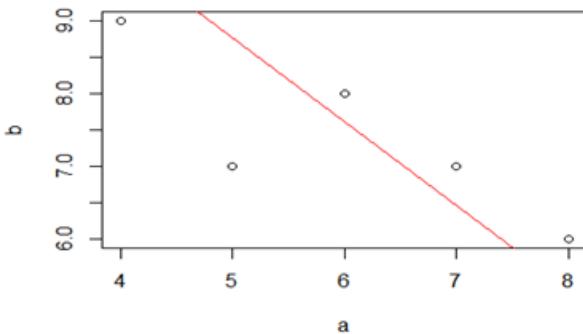
```
b<-c(9,7,8,7,6)
```

```
cor(a,b)
```

```
## [1] -0.8320503
```

En este caso hay una correlación negativa fuerte. También se pueden construir regresiones, como la regresión lineal y graficas como la gráfica de dispersión.

```
linearMod <- lm(a~b)  
plot(x=a, y=b)  
abline(lm(a~b), col="red")
```

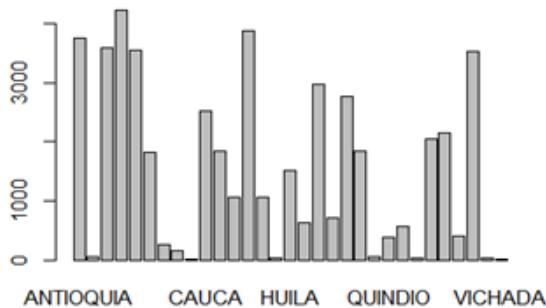


Actividad 7: Escoja una columna de datos cuantitativos del dataframe que seleccionó. Calcule la media, mediana, varianza y desviación estándar de los datos en esa columna. Si su dataframe tiene dos columnas numéricas, realice un diagrama de dispersión y una recta de regresión.

También se pueden construir otras graficas básicas, como los gráficos de barras, histogramas y boxplot.

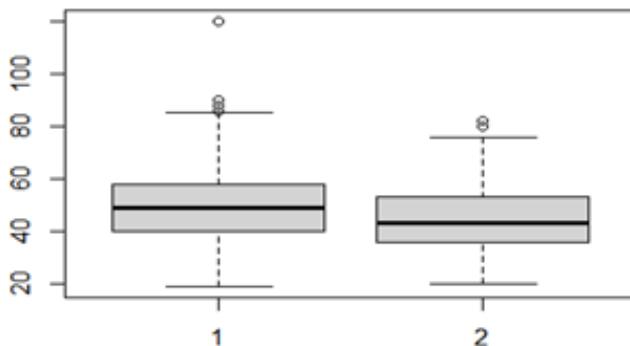
Por ejemplo, para calcular el grafico de barras con la cantidad de madres y padres comunitarios por departamento

```
barplot(table(Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF$DEPARTAMENTOUDS))
```



Si se quiere un boxplot comparativo entre las edades de las madres y padres comunitarios de Valle y Cundinamarca

```
boxplot(Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF[Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF$DEPARTAMENTOUDS=='VALLE DEL CAUCA', ]$EDADENAÑOS, Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF[Caracterizaci_n_Madres_y_Padres_Comunitarios_ICBF$DEPARTAMENTOUDS=='CUNDINAMARCA', ]$EDADENAÑOS)
```



Vale la pena anotar que los gráficos obtenidos son muy sencillos visualmente. R puede modificarles los colores y esquemas y agregar estilos.

Actividad 8: Realice una gráfica para una de las variables cualitativas y una de las variables cuantitativas de su dataframe.

Referencias

- Datos Abiertos Colombia. (s. f.). Caracterización Madres y Padres Comunitarios ICBF. Recuperado 27 de marzo de 2021, de
<https://www.datos.gov.co/dataset/Caracterizaci-n-Madres-y-Padres-Comunitarios-ICBF/ixwb-p9qb>
- R. (s. f.). CRAN - Mirrors. Recuperado 27 de marzo de 2021, de
<https://cran.r-project.org/mirrors.html>
- RStudio. (s. f.). Download the RStudio IDE. Recuperado 27 de marzo de 2021, de
<https://rstudio.com/products/rstudio/download/>
- R. (s. f.). The R Project for Statistical Computing. Recuperado 27 de marzo de 2021, de <https://www.r-project.org/>
- Santana, J. S., & Farfán, E. M. (s. f.). El arte de programar en R: un lenguaje para la estadística. 197.



¡ideas que crean valor!