

SESIÓN 01:

Fundamentos de Programación en R

DOCENTE: LUIS FELIPE GARAYAR BURNEO

CORREO: LUIS.GARAYAR.DMC@GMAIL.COM









Reglas



Puntualidad



Mantener silenciado el micrófono durante la sesión



Las preguntas se realizarán por el chat/ en caso sea necesario se habilita el micrófono



Realizar las actividades encomendadas



SOBRE EL DOCENTE

- ✓ Ingeniero Estadístico UNI
- ✓ Ingeniero de Sistemas UNMSM
- ✓ Candidato a Magister en Estadística Aplicada UNALM
- ✓ Data Scientist Practicioner UNIR (España)
- ✓ Más de 10 años de experiencia en Banca (Inteligencia Comercial: Campañas Dirigidas, Reportería, Análisis y Modelos Estadísticos)
- ✓ Actualmente laborando en el BCP.
- ✓ Campos de Interés: Excel, Macros (VBA), Dashboards en Excel, Power BI, Lenguaje SQL, Oracle PL/SQL, software estadístico R y Python





Calificación

| Asistencia (Curso) | (10%) |
|--------------------|-------|
|--------------------|-------|

Trabajos calificados (40%)

Examen Final (teoría y práctica) (50%)







- I. Introducción al R y al R Studio
- II. Tipos y estructuras de datos en R
- III. Programación básica en R
- IV. Análisis Gráfico con R
- V. Manipulación Avanzada de Dataframes
- VI. Introducción al Análisis Estadístico con R
- VII. Publicación de Resultados







FECHA DE ENTREGAS DE TRABAJOS

• Listado de Ejercicios Sesión 1 Sábado 26/09 medianoche.

• Listado de Ejercicios Sesión 2 Miércoles 30/09 medianoche.

• Listado de Ejercicios Sesión 3 Sábado 03/10 medianoche.

• Listado de Ejercicios Sesión 4 Jueves 08/10 medianoche.

• Listado de Ejercicios Sesión 5 Domingo 11/10 medianoche.

• Examen Final Teórico y Práctico

Formulario Activo entre Martes 06/10 y Domingo 11/10.



I. Introducción al R y al R Studio

- II. Tipos y estructuras de datos en R
- III. Programación básica en R
- IV. Análisis Gráfico con R
- V. Manipulación Avanzada de Dataframes
- VI. Introducción al Análisis Estadístico con R
- VII. Publicación de Resultados







Un poco de historia...



- R es básicamente un lenguaje de programación que permite realizar comandos e implementar técnicas estadísticas en un entorno interactivo para el análisis estadístico y gráfico.
- Es un lenguaje de programación con funciones orientadas a objetos.
- R fue inicialmente diseñado por Robert Gentleman y Ross Ihaka (1993), miembros del Departamento de Estadística de la Universidad de Auckland, en Nueva Zelanda.



















• La sintaxis es simple e intuitiva.

```
#Creando vectores.

> x<-c(10.4,5.6,3.1,6.4,21.7)
> x

[1] 10.4 5.6 3.1 6.4 21.7

> w<-c("rojo","verde","azul")
> w

[1] "rojo" "verde" "azul"
```

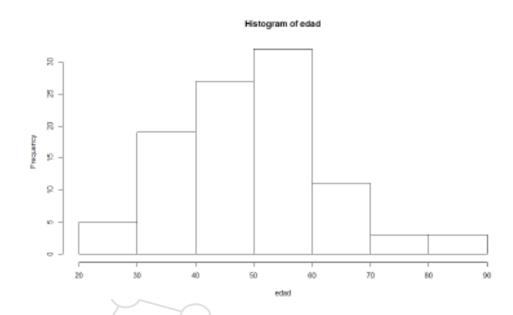
```
#Creando objetos y realizando cálculos aritméticos.
> a=10
> b=25
> c=25
> y=a+b+c
> y
[1] 60
```

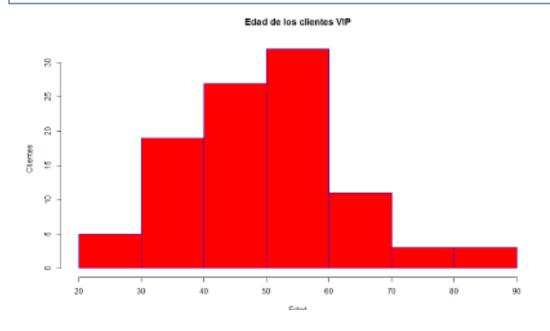


 La estructura y facilidad de uso de R nos permite implementar nuestras propias funciones y rutinas a medida que aparecen nuestras necesidades.



>hist(edad,col="red", xlab="Edad",ylab="Clientes",
main="Edad de los clientes VIP",border="blue")

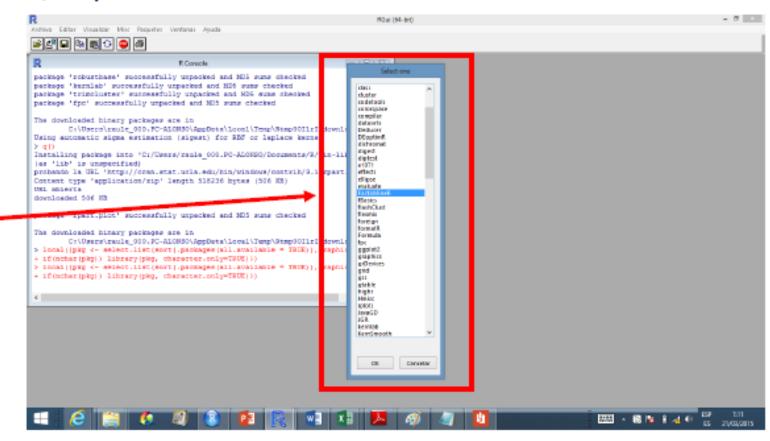






 La distribución de R viene acompañada de un numeroso conjunto de librerías base. Así mismo, es posible añadir librerías adicionales.

Librerías propias y librerías adicionales como el FactoMineR











• Gran variedad de librerías gráficas.





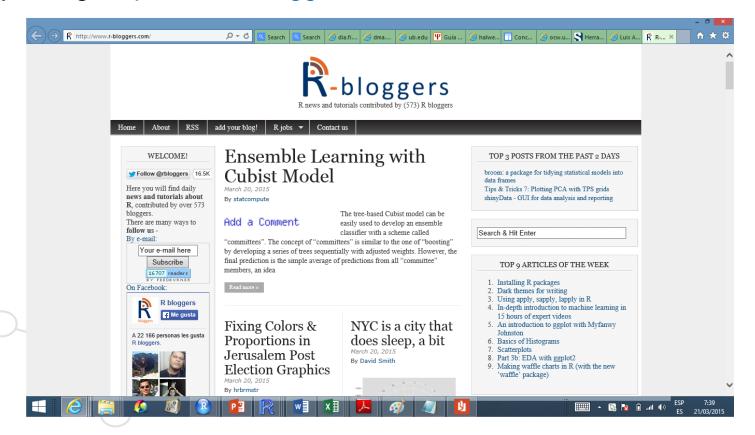




Gran red de apoyo y soporte disponible en foros, blogs, Facebook, etc.

El mejor foro: http://stackoverflow.com/

El mejor blog: http://www.r-bloggers.com/



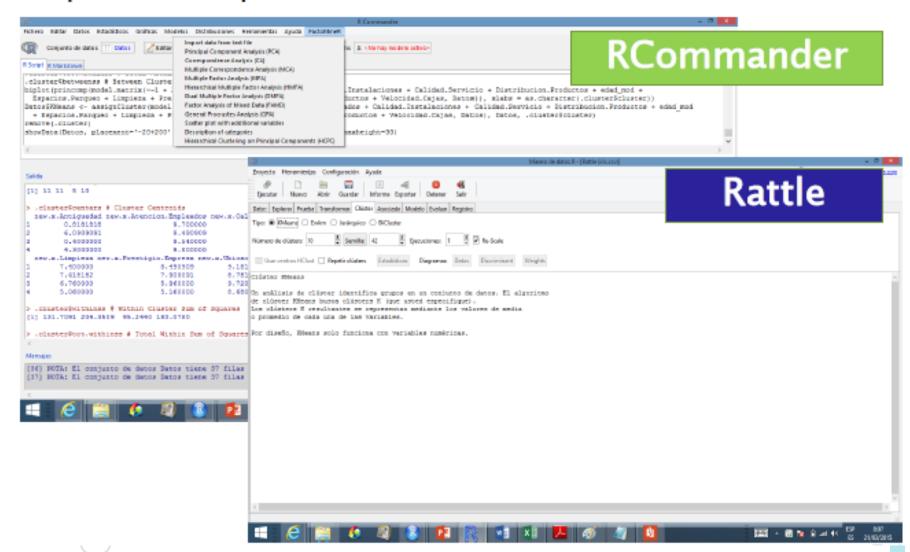








Paquetes de R para la Minería de Datos.





Entorno R y R Studio



Podemos definir R desde dos perspectivas distintas:

R es un entorno software

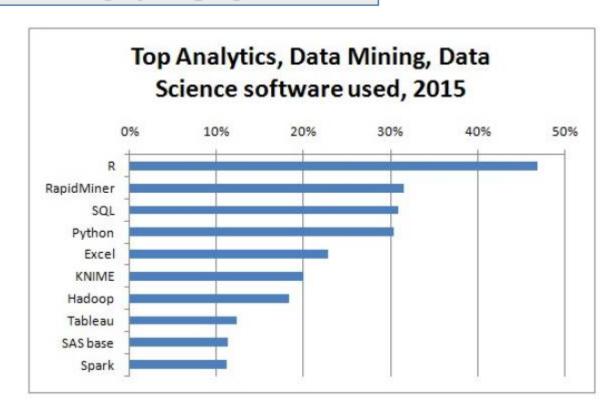
R es un lenguaje de programación

R Permite:

- > Cargar y Manipular datos.
- Realizar Análisis.
- Presentar los resultados gráficamente

Ventajas de R

- Herramienta libre y gratuita(GNUGPL).
- Evolución continua y rápida.
- Versiones en múltiples sistemas operativos.
- > Flexible y potente.
- Mucha documentación/ayuda. Multitud de librerías y extensiones.





Entorno R y R Studio

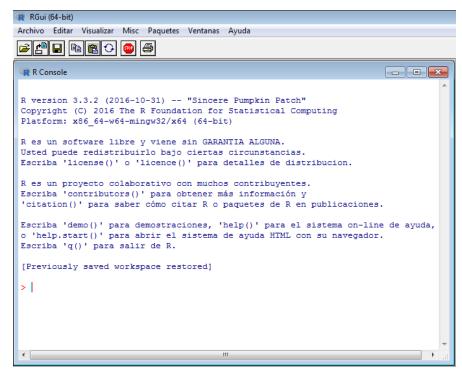




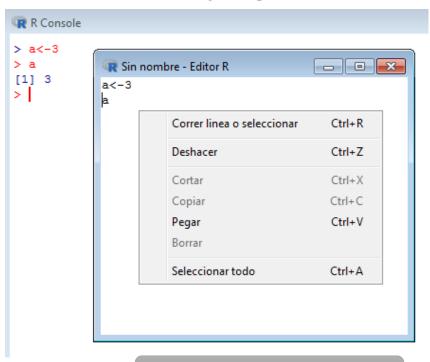
Entorno R



Consola R



Editor de programas



Atajo Ejecutar : CTRL+ R

- *Prompt (>):* Indica que R está esperando comandos
- Se ingresa comandos y se ejecuta en consola con un salto de línea.
- Otra forma de ingresar comandos es mediante un Editor de scripts R.
- En este caso, se ejecuta por línea completa (donde se encuentra el cursor) o una selección (puede ser parte de una línea o seleccionando varias líneas).

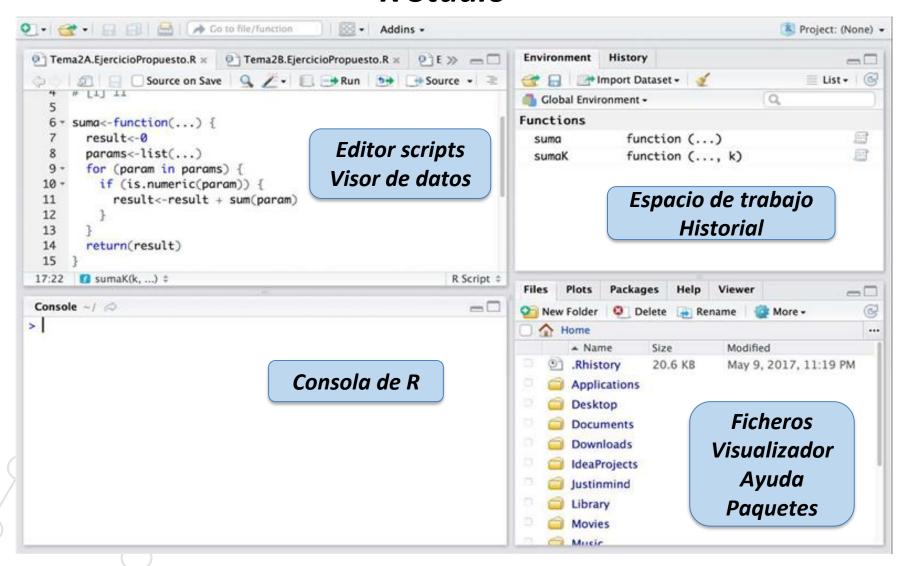


Entorno R Studio



R Studio

Atajo Ejecutar : *CTRL + Enter*

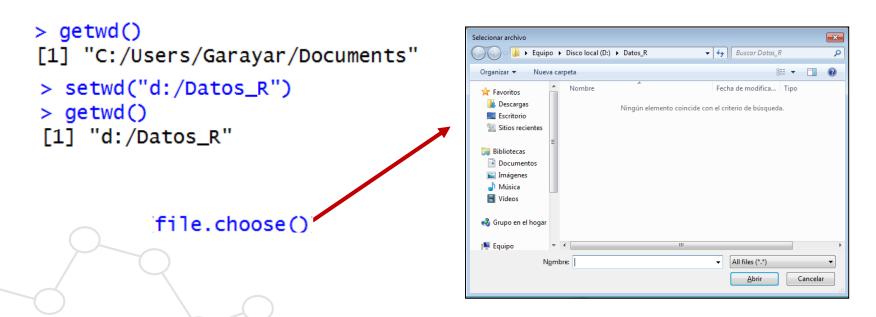






Workspace (espacio de trabajo)

- Es el lugar donde se almacena todo lo que se realiza en una sesión de R.
- ➤ Para obtener la dirección actual del Workspace se usa la función *getwd*.
- ➤ Para cambiar la ruta que usaremos para el Workspace, usaremos setwd.
- En general, si no tenemos la ruta de la carpeta que seleccionaremos, podemos ubicarla con *file.choose()*, la que nos abrirá un navegador de archivos.







Objetos en R

 Son todas las entidades que manipula R. Por ejemplo, vectores de números, vectores lógicos, vectores de caracteres, matrices, etc.

• Los atributos de un objeto suministran información específica sobre el propio

objeto.

| Atributo | Descripción | |
|-------------|--|--|
| Modo | Cualquier tipo de entidad que maneja R. Se usa la función: mode(objeto). | |
| Tipo | Tipo de datos de los objetos: entero, carácter, double, etc. Se usa la función: typeof(objeto). | |
| Nombre | Etiquetas de los elementos individuales de un vector o lista. Se usa la función: names(objeto). | |
| Dimensiones | Dimensiones de las matrices (alguna puede ser cero). Se usa la función: dim(objeto). | |
| Dimnames | Nombres de las dimensiones de la matriz. Se usa la función: dimnames(objeto). | |
| Clase | Vector alfanumérico con la lista de las clases del objeto. Se usa la función: class(objeto). | |





Creación de Objetos y Variables

- ➤ Se usa el operador de asignación (<-)
- ➤ También se puede usar el operador (=)

Reglas para la creación de nombres de variables







R distingue entre mayúsculas y minúsculas. Usualmente se permite utilizar toda clase de símbolos alfanuméricos (incluyendo caracteres acentuados, ñ, etc.) además del punto «.» y del guion bajo «_». Eso sí, los nombres deben empezar por «.» o por una letra y si empiezan por «.» entonces el segundo carácter no debería ser un número..

El entorno no impone límites a la longitud de los nombres de los símbolos (aunque por motivos prácticos no deberíamos hacerlos demasiado largos).

Ejemplos:

data_año1 Data_año1 .var9 tabla.final..piloto.PLD_2



Algunas consideraciones:

- ➤ Para obtener el valor de una variable u objeto por consola, tecleamos su nombre seguido de un fin de línea. También se puede usar la función **print**.
- ➤ Para modificar el valor de un objeto o variable simplemente lo reasignamos.
- ➤ Los caracteres se colocan entre comillas simples (' ') o dobles (" ").
- ➤ Se pueden realizar varios comandos en una línea separados por punto y coma (;)

```
> name <- "Carmen"; n1 <- 10; n2 <- 100; m <- 0.5
> n1
[1] 10
> name
[1] "Carmen"
> name <- 'Luis Felipe'
> name
[1] "Luis Felipe"
> print(name)
[1] "Luis Felipe"
```







➤ La función [s] nos permite listar los objetos existentes en memoria.

También se puede usar la función objects.

La opción **pattern (pat)** nos permite realizar búsquedas entre las variables.

➤ La función *ls.str* nos da algunos detalles de las variables y objetos en memoria.

➤ La función <u>rm</u> borra objetos de la memoria.

```
> 1s()
                          "name"
> objects()
                          "name"
> ls(pat = "m") #que contenga m
           "name"
> ls(pat = "^m") #que empiece con m
[1]
> ls.str()
     num 0.5
      num 100
name : chr "Luis Felipe"
> rm(name)
> 1s()
```







[1] 3.14159265358979



 R almacena los valores con el máximo número de dígitos, hasta 17, pero muestra solo 7, si uno desea que muestre mayor número, debemos cambiar esta opción.
 La función options determina como R calcula y muestra los resultados.

```
> pi  # R por defecto muestra 7 dígitos
[1] 3.141593
> options(digits=15) # Cambia a 15 dígitos
```

>En consola, se puede navegar en comandos anteriores con las teclas direccionales.

En cualquier momento se puede guardar la sesión con **save.image()** (si no se especifica ruta, se guardará en la ruta definida para Workspace)

```
Limpiar consola: CTRL + L
```

> pi

Salir de la sesión: q()

```
> q()
Save workspace image to ~/.RData? [y/n]: y
```

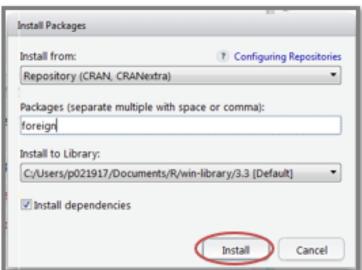


Instalación y carga de paquetes



- Una ventaja de R es que es fácilmente expandible usando paquetes.
- Existen más de 8000 paquetes actualmente.
- Estos paquetes se obtienen de servidores CRAN esparcidos geográficamente.
- ➤ En consola, instalamos un paquete con install.packages
 - > install.packages("foreign", dependencies=TRUE)
- Esta sentencia instala el paquete "foreign" con otros paquetes que dependen de este. Luego cargamos el paquete para su uso: > library("foreign")
- ➤ En R Studio hay otra forma de instalar paquetes:





https://cran.r-project.org/web/views/



- Las vistas de tareas de CRAN tienen como objetivo proporcionar una guía sobre qué paquetes en CRAN son relevantes para las tareas relacionadas con un tema determinado.
- Ofrecen una breve descripción general de los paquetes incluidos.
- Las vistas están destinados a tener un enfoque claro para qué paquetes deben ser incluidos (o excluidos) y son *no* destinado a apoyar los "mejores" paquetes para una tarea determinada.

Bayesian Inference

<u>ChemPhys</u> Chemometrics and Computational Physics

Clinical Trial Design, Monitoring, and Analysis

<u>Cluster</u> Cluster Analysis & Finite Mixture Models

<u>DifferentialEquations</u>
Distributions
Distributions
Differential Equations
Probability Distributions

<u>Econometrics</u> Econometrics

Environmetrics Analysis of Ecological and Environmental Data

Experimental Design of Experiments (DoE) & Analysis of Experimental Data

Extreme Value Analysis

<u>Finance</u> Empirical Finance

<u>FunctionalData</u> Functional Data Analysis

Genetics Statistical Genetics

Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization

HighPerformanceComputing High-Performance and Parallel Computing with R

Machine Learning & Statistical Learning

Medical Image Analysis

Meta-Analysis Meta-Analysis

<u>ModelDeployment</u> Model Deployment with R <u>Multivariate</u> Multivariate Statistics

Natural Language Processing Natural Language Processing

Numerical Mathematics Numerical Mathematics

Official Statistics & Survey Methodology
Optimization Optimization and Mathematical Programming

Pharmacokinetics Analysis of Pharmacokinetic Data

<u>Phylogenetics</u> Phylogenetics, Especially Comparative Methods

<u>Psychometrics</u> Psychometric Models and Methods

Reproducible Research
Robust
Robust
Reproducible Research
Robust
Robust Statistical Methods

Social Sciences Statistics for the Social Sciences

Spatial Analysis of Spatial Data

<u>SpatioTemporal</u> Handling and Analyzing Spatio-Temporal Data

<u>Survival</u> Survival Analysis TimeSeries Time Series Analysis

WebTechnologies Web Technologies and Services

gRaphical Models in R



Expresión

- Se evalúa, se muestra y el valor se pierde
- Ejemplo: 10 * 3

Asignación

- Evalúa la expresión, no se muestra y guarda el resultado en una variable u objeto
- Ejemplo: x <- 10 * 3
- Operaciones matemáticas +, -, *, /, ^
- Operadores de comparación <, ==, >, <=, >=, !=
- Operadores lógicos (and, or, not) &, |, !





Funciones para clases de datos

- class()
- is.classname()
 - is.numeric()

is.integer()

is.character()

- as.classname()
 - as.numeric()

as.integer()

as.character()



Operaciones con valores lógicos TRUE FALSE

• ¿Es una clase lógica? is.logical()

Conectiva lógica AND:

Conectiva lógica OR:

Conectiva lógica NOT:







Realizando operaciones básicas

> 2+3

[1] 5

> 2*3

[1] 6

> 2/3

[1] 0.6666667

> 2^3

[1] 8

> 2**3

[1] 8

> 3%%2

[1] 1

> (2.2+3.5)*2.1

[1] 11.97

> 1+2+3-4

[1] 2

> 5%/%2

[1] 2

Llamando a funciones

> exp(1)

[1] 2.718282

> log(exp(1))

[1] 1

> log10(10)

[1] 1

> log2(2)

[1] 1

> sqrt(4)

[1] 2

> cos(pi)

[1] -1

> sin(pi)

[1] 1.224647e-16

> tan(pi)

[1] -1.224647e-16

> abs(-2.4)

[1] 2.4

> ceiling(-2.4)

[1] -2

> floor(-2.4)

[1] -3

> round(-2.3894, 1)

[1] -2.4

> round(-2.3894, digits=1)

[1] -2.4

> round(digits=1, x=-2.3894)

[1] -2.4







| Operador | Definición | Ejemplo |
|---------------|--|---|
| abs(a) | Valor absoluto de a | > abs(-4) [1].4 |
| sign(a) | Devuelve el signo de a: 1 E Positivo 0 E Cero -1 E Negativo | > sign(-4) [1]1 > sign(11-11) [1].0 > sign(8) [1].1 |
| round(a, dec) | Redondea un número según el número de decimales | > round(13.68,1) [1] 13.7 |
| sqrt(a) | Raíz cuadrada de a | > sqrt(4) [1].2 |
| log(a) | Logaritmo exponencial | > log(4) [1].1.386294 |
| log10(a) | Logaritmo de a de base 10 > log10(10) [1].1 | |
| exp(a) | Función exponencial de a > exp(1) [1].2.718282 | |

Uso de la ayuda en R



La función **help** nos permite acceder a la ayuda en R. Es igual si se antecede ? a la función.

➤ Si se busca ayuda sobre una función especial (como un operador matemático) se busca con comillas (simples o dobles)

```
> help(mean)
> help("mean")
> ?mean

> ?*
Error: unexpected '*' in "?*"
> ?'*'
> help("*")
```

La función **help.search** busca toda la ayuda que exista con una palabra clave entre comillas (también se puede usar ??)

```
> ??mean
> help.search("mean")
```







I. Introducción al R y al R Studio

II. Tipos y estructuras de datos en R

- III. Programación básica en R
- IV. Análisis Gráfico con R
- V. Manipulación Avanzada de Dataframes
- VI. Introducción al Análisis Estadístico con R
- VII. Publicación de Resultados

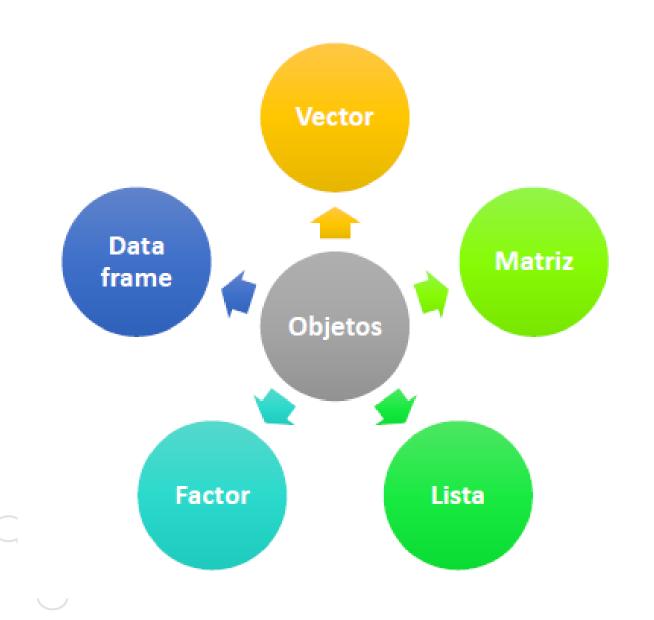






Estructuras de datos básicos en R

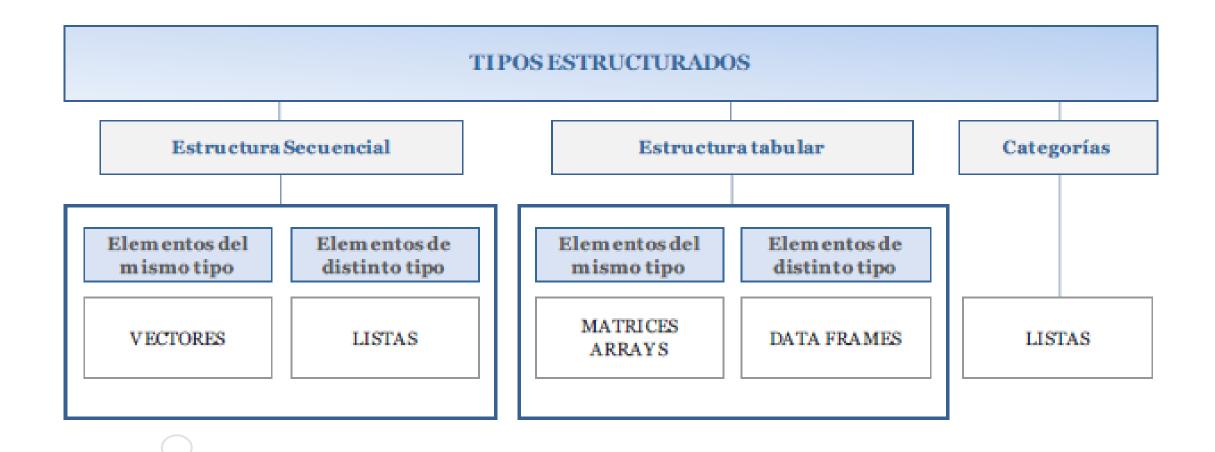






Estructuras de datos básicos en R





Estructuras de datos en R: Vectores



Estructura de datos: vectores

- Secuencias de elementos del mismo tipo
- Se crean con:
 - Función c
 - Operador :
 - Función seq
 - Función rep
 - Función append

```
R Console
                                                    Q Help Search
> c(1,2.3,-0.56)
[1] 1.00 2.30 -0.56
> 1:5
[1] 1 2 3 4 5
> seq(-1,1,0.5)
[1] -1.0 -0.5 0.0 0.5 1.0
rep(c(1,2),3)
[1] 1 2 1 2 1 2
> c(c(1.2,2.7),3:5)
[1] 1.2 2.7 3.0 4.0 5.0
> v<-1:5</p>
> append(v, c(6,7))
[1] TRUE FALSE FALSE TRUE
> c("Alma", "Casa", "Marte")
[1] "Alma" "Casa" "Marte"
```









 Un vector es una colección ordenada de elementos del mismo tipo.

Utilizamos la función c para generar vectores.

```
#Creando vectores.

> x<-c(10.4,5.6,3.1,6.4,21.7)
> x
[1] 10.4 5.6 3.1 6.4 21.7

> y<-c(1:5)
> y
[1] 1 2 3 4 5

> w<-c("rojo","verde","azul")
> w
[1] "rojo" "verde" "azul"
```



Estructuras de datos en R: Vectores



Generación de sucesiones

```
> 1:20
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
> 1:3*5+2#Se realiza la sucesión primero y luego las operaciones aritméticas
[1] 7 12 17
> seq(1,10,by=2)#Genera una secuencia que inicia en 1 y termina en 10 con elementos que van de 2 en 2
[1] 1 3 5 7 9
```

- Los vectores lógicos aparecen en condiciones lógicas.
- Sólo pueden tomar dos valores TRUE(verdadero) y FALSE(falso).

```
> temp<-x>7
> temp
[1] TRUE FALSE FALSE FALSE TRUE
```





Vectores índices y sub-vectores



- Para seleccionar un sub-vector o un elemento de un vector se utiliza un vector de índices.
- Estructura: nombre del vector[vector de índices]

```
> x
[1] 10.4 5.6 3.1 6.4 21.7
> x[4]#Devuelve el elemento de la posición 4
[1] 6.4
> x[1:3]#Devuelve los elementos de x de las posiciones 1,2,3
[1] 10.4 5.6 3.1
> x[-(3:4)]#Devuelve los elementos de x sin considerar los elementos de la posición 3 y 4
[1] 10.4 5.6 21.7
```





Funciones con Vectores



- Operadores aritméticos elementales +,-,*,/ y para la potencia ^.
- Las operaciones se efectúan elemento a elemento.

```
> v<-4*x-y+2
> v
[1] 34.2 20.8 9.3 20.2 66.1
```

Funciones conocidas como log(), exp(), sin(), cos(), tan(), sqrt(), max(), min(), lenght(), mean(), var(x), sd(), sort().





Funciones con Vectores



| sum(x) | prod(x) | Sum/product of the elements of x | |
|-----------------|--------------|--|--|
| cumsum(x) | cumprod(x) | Cumulative sum/product of the elements of | |
| min(x) | max(x) | Minimum/Maximum element of x | |
| mean(x) | median(x) | Mean/median of x | |
| var(x) | sd(x) | Variance/standard deviation of x | |
| cov(x,y) | cor(x,y) | Covariance/correlation of x and y | |
| range(x) | | Range of x | |
| quantile(x) | | Quantiles of x for the given probabilities | |
| fivenum(x) | | Five number summary of x | |
| length(x) | | Number of elements in x | |
| unique(x) | | Unique elements of x | |
| rev(x) | | Reverse the elements of x | |
| sort(x) | | Sort the elements of x | |
| which() | | Indices of TRUEs in a logical vector | |
| which.max(x) | which.min(x) |) Index of the max/min element of x | |
| match() | | First position of an element in a vector | |
| union(x, y) | | Union of x and y | |
| intersect(x, y) | | Intersection of x and y | |
| setdiff(x, y) | | Elements of x that are not in y | |
| setequal(x, y) | | Do x and y contain the same elements? | |





· GRACIAS!

