

# R Sintaxis: Un resumen rápido

Mario Gavidia-Calderón

3/27/2021

# R: Sintaxis

- ▶ R como calculadora:

```
(5 + 10 * 2 / 4) ^ 2 - 5
```

```
## [1] 95
```

- ▶ Para **asignar** valor a una variable <-

```
R <- 8.314
```

```
R
```

```
## [1] 8.314
```

## R: Sintaxis

- Para comentar #

```
R <- 8.314 # Constante general de los gases (J K / mol)
R
```

```
## [1] 8.314
```

- Una función se ejecuta así: `nombre_de_la_funcion()`

```
class(R)
```

```
## [1] "numeric"
```

# R: Objetos

## ► character

```
una_palabra <- "palabra"  
class(una_palabra)
```

```
## [1] "character"
```

## ► numeric

```
this_year <- 2021  
g <- 9.81 # m/s2  
class(this_year)
```

```
## [1] "numeric"
```

```
class(g)
```

```
## [1] "numeric"
```

# R: Objetos

## ► booleans

```
verdad <- TRUE  
verdad
```

```
## [1] TRUE
```

```
falso <- 5 > 10  
falso
```

```
## [1] FALSE
```

```
falsazo <- "cinco" == "5"  
falsazo
```

```
## [1] FALSE
```

## R: Objetos - Vectores

- un vector se define usando `c()`

```
puntos_cardinales <- c("N", "E", "S", "W")  
puntos_cardinales
```

```
## [1] "N" "E" "S" "W"
```

```
puntos_cardinales_grados <- c(0, 90, 180, 270)  
class(puntos_cardinales_grados)
```

```
## [1] "numeric"
```

## R: Objetos - Vectores

- Una secuencia usando `seq(inicio, final, intervalo)`

```
del_1a15 <- seq(1, 5)
del_1a15
```

```
## [1] 1 2 3 4 5
```

```
pares_hasta10 <- seq(0, 10, 2)
pares_hasta10
```

```
## [1] 0 2 4 6 8 10
```

```
sec_float <- seq(0, 1, 0.2)
sec_float
```

```
## [1] 0.0 0.2 0.4 0.6 0.8 1.0
```

## R:Objetos - Vectores - Selección de elementos

- ▶ Para seleccionar elementos de un vector:  
`nombre_vector[posición]:`

```
# Primer elemento
```

```
puntos_cardinales_grados[1]
```

```
## [1] 0
```

```
# Último elemento
```

```
puntos_cardinales_grados[4]
```

```
## [1] 270
```



## R:Objetos - Vectores - Selección de elementos

- Pueden ser varios elementos usando otro vector

```
# Segundo y tercero  
puntos_cardinales[c(2, 3)]
```

```
## [1] "E" "S"
```

- Se puede eliminar elementos usando  
nombre\_vector[-posición]

```
GEE <- c("H2O", "CO2", "O2", "CH4")  
GEE
```

```
## [1] "H2O" "CO2" "O2"  "CH4"
```

```
# Oxígeno no es GEE  
GEE[-3]
```

```
## [1] "H2O" "CO2" "CH4"
```

## R:Objetos - Vectores - Substitución

- Podemos substituir un elemento de un vector así:

```
# Reemplazamos Oxígeno por Ozone
```

```
GEE[3] <- "O3"
```

```
GEE
```

```
## [1] "H2O" "CO2" "O3"  "CH4"
```

## R: Objetos - data frames

- ▶ Un data frame es una **tabla**
- ▶ Una matriz **indexada**: nombres de columnas y filas
- ▶ Cada **columna** sería una **variable**
- ▶ Cada **fila** es una **observación**
- ▶ Un conjunto de vectores

## R: Objetos - data frame

- Creamos un data frame con la función `data.frame()`

```
gases <- c("N2", "O2", "Ar", "CO2")
peso_molecular <- c(28, 32, 40, 12 + 2 * 16)
porcentaje <- c(78.08, 20.95, 0.9, 0.04)
```

```
aire <- data.frame(gas = gases,
                   W = peso_molecular,
                   per = porcentaje)
```

```
aire
```

```
##   gas  W  per
## 1  N2 28 78.08
## 2  O2 32 20.95
## 3  Ar 40  0.90
## 4 CO2 44  0.04
```

## R: Objetos - data frame

- Creamos un data frame con la función `data.frame()`

*# O directamente*

```
aire <- data.frame(gas = c("N2", "O2", "Ar", "CO2"),  
                  W = c(28, 32, 40, 12 + 2 * 16),  
                  per = c(78.08, 20.95, 0.9, 0.04))
```

aire

```
##   gas  W  per  
## 1  N2 28 78.08  
## 2  O2 32 20.95  
## 3  Ar 40  0.90  
## 4 CO2 44  0.04
```

## R: Objetos - data frame

Un ejemplo con los Beatles

```
beatles <- data.frame(  
  name = c("John", "Paul", "George", "Ringo"),  
  born = c(1940, 1941, 1943, 1940),  
  inst = c("guitar", "bass", "guitar", "drums"),  
  stringsAsFactors = F)  
beatles
```

```
##      name born  inst  
## 1   John 1940 guitar  
## 2   Paul 1941  bass  
## 3 George 1943 guitar  
## 4  Ringo 1940 drums
```

## R: data frame - Seleccionando valores

beatles[fila, columna] - fila y columna son índices de posición:  
**enteros**

```
beatles
```

```
##      name born  inst
## 1   John 1940 guitar
## 2   Paul 1941  bass
## 3 George 1943 guitar
## 4  Ringo 1940  drums
```

```
beatles[1, 1]
```

```
## [1] "John"
```

## R: data frame - Seleccionando valores

- ▶ Seleccionando a Paul, y su año de nacimiento

```
beatles[2, c(1, 2)]
```

```
##   name born  
## 2 Paul 1941
```

- ▶ Seleccionando a los guitarristas

```
##      name born  inst  
## 1   John 1940 guitar  
## 3 George 1943 guitar
```



## R: data frame - Seleccionando valores

- ▶ Espacio en blanco = “**Todo**”
- ▶ *Todo* los nombres

```
beatles[, 1]
```

```
## [1] "John"    "Paul"    "George"  "Ringo"
```

- ▶ **Todos** los datos de Ringo

```
##      name born  inst  
## 4 Ringo 1940 drums
```

## R: data frame - Seleccionando valores

- ▶ Una columna = Un sólo índice
- ▶ *Todos* los instrumentos

```
beatles[3]
```

```
##      inst  
## 1 guitar  
## 2  bass  
## 3 guitar  
## 4 drums
```

```
class(beatles[3])
```

```
## [1] "data.frame"
```

## R: data frame - \$ (signo de dolar)

- ▶ Seleccionamos una **columna** de un **data frame** como un **vector**
- ▶ Sintaxis: `df$nombre_columna`
- ▶ E.g. Nombre de los Beatles

```
beatles$name
```

```
## [1] "John"    "Paul"    "George"  "Ringo"
```

```
class(beatles$name)
```

```
## [1] "character"
```

- ▶ **Ojo!**: `beatles[1] ≠ beatles$name`

## R: data frame - \$ (signo de dolar)

- ▶ Algunas funciones necesitan vectores como **input**
- ▶ Média de año de nacimiento

```
mean(beatles[2])
```

```
## Warning in mean.default(beatles[2]): argument is not numeric
## returning NA

## [1] NA
```

```
mean(beatles$born)
```

```
## [1] 1941
```

## R: data frame - Creando nuevas columnas

- ▶ Usamos \$ así: `df$nueva_columna <-`
- ▶ Mejores canciones como solistas (según yo)

```
beatles$solo <- c("Imagine",  
                  "Band on the run",  
                  "Oh sweet Lord",  
                  "Photograph")
```

```
beatles
```

##	name	born	inst	solo
## 1	John	1940	guitar	Imagine
## 2	Paul	1941	bass	Band on the run
## 3	George	1943	guitar	Oh sweet Lord
## 4	Ringo	1940	drums	Photograph

## R: data frame - Algunas funciones

- ▶ Número de filas: `nrow()`
- ▶ Número de columnas: `ncol()`

```
nrow(beatles)
```

```
## [1] 4
```

```
ncol(beatles)
```

```
## [1] 4
```

## R: data frame - Algunas funciones

- Tipo de objeto de cada columna: `str()`

```
str(beatles)
```

```
## 'data.frame':    4 obs. of  4 variables:
##  $ name: chr  "John" "Paul" "George" "Ringo"
##  $ born: num  1940 1941 1943 1940
##  $ inst: chr  "guitar" "bass" "guitar" "drums"
##  $ solo: chr  "Imagine" "Band on the run" "Oh sweet Lord"
```

- nombre de columnas

```
names(beatles)
```

```
## [1] "name" "born" "inst" "solo"
```

## R: data frame - Algunas funciones

- ▶ Valores únicos: `unique()`

```
# Instrumentos
```

```
unique(beatles$inst)
```

```
## [1] "guitar" "bass"   "drums"
```

```
# Año nacimiento
```

```
unique(beatles$born)
```

```
## [1] 1940 1941 1943
```



## R: data.frame - Algunas funciones

- ▶ Primeras observaciones: `head()`
- ▶ últimas observaciones: `tail()`

```
head(beatles)
```

```
##      name born  inst      solo
## 1   John 1940 guitar    Imagine
## 2   Paul 1941  bass Band on the run
## 3 George 1943 guitar Oh sweet Lord
## 4  Ringo 1940 drums    Photograph
```

```
tail(beatles)
```

```
##      name born  inst      solo
## 1   John 1940 guitar    Imagine
## 2   Paul 1941  bass Band on the run
## 3 George 1943 guitar Oh sweet Lord
## 4  Ringo 1940 drums    Photograph
```

## R: data.frame - Algunas funciones

- ▶ Primeras observaciones: `head()`
- ▶ últimas observaciones: `tail()`

```
head(beatles, 2)
```

```
##   name born  inst      solo  
## 1 John 1940 guitar  Imagine  
## 2 Paul 1941  bass Band on the run
```

```
tail(beatles, 2)
```

```
##      name born  inst      solo  
## 3 George 1943 guitar Oh sweet Lord  
## 4  Ringo 1940  drums  Photograph
```

## R: data.frame - Substitución

```
# Otras canciones
```

```
beatles$solo <- c("Working class heroes",  
                 "Maybe I'm amazed",  
                 "What is life?",  
                 "I don't come easy")
```

```
beatles
```

```
##      name born  inst      solo  
## 1   John 1940 guitar Working class heroes  
## 2   Paul 1941  bass   Maybe I'm amazed  
## 3 George 1943 guitar   What is life?  
## 4  Ringo 1940 drums    I don't come easy
```

## R: Operaciones *Element-wise*

- ▶ Cuántos años tendrían los Beatles?

```
2021 - beatles$born
```

```
## [1] 81 80 78 81
```

```
print(beatles$name)
```

```
## [1] "John"    "Paul"    "George"  "Ringo"
```

## R: Operaciones *Element-wise*

```
tempC <- c(27, 32, 28, 26)
tempK <- tempC + 273.15
tempK
```

```
## [1] 300.15 305.15 301.15 299.15
```

```
tempk_chr <- as.character(tempK)
str(tempk_chr)
```

```
## chr [1:4] "300.15" "305.15" "301.15" "299.15"
```

## De Excel a R: `read.table()`

- ▶ Usar Excel para exportar la hoja de cálculo como `.csv`
- ▶ Evitar caracteres especiales en los nombres de las columnas:
  - ▶ No tildes
  - ▶ No espacios
  - ▶ No paréntesis, etc
- ▶ Usar nombres más sencillos:
  - ▶ **tempC** en vez **Temperatura en (°C)**
- ▶ Para importar una tabla a R usamos `read.table()`

## De Excel a R: Vamos a leer la encuesta

```
survey <- read.table(  
  "../02_data/respuestas27.csv", # Dónde está el archivo  
  header = T, # Las columnas tienen nombre?  
  sep = ",", # Cómo están separadas las columnas  
  stringsAsFactors = F # caracteres no son factor  
)  
head(survey[, -c(2, 3)]) # No nombres apellidos
```

```
##           Timestamp y.tu.edad.           Vivo.en. P  
## 1 3/25/2021 1:32:34           24 Villa María del Triunfo  
## 2 3/25/2021 1:32:47           27 Fuera de Lima La Gris  
## 3 3/25/2021 1:36:09           27 Fuera de Lima La Gris  
## 4 3/25/2021 1:45:37           26 San Martín de Porres  
## 5 3/25/2021 1:52:05           23 San Juan de Miraflores  
## 6 3/25/2021 2:54:47           25 Villa El Salvador  
## Qué.estudias.estudiaste. En.qué.año.vas. Sabes.programar  
## 1 Ingeniería forestal           5  
## 2 Ingeniería ambiental Ya me gradué  
## 3 Ingeniería ambiental Ya me gradué
```

# De Excel a R: Vamos a leer la encuesta

## ► Mejores nombres de columnas

```
names(survey) <- c("date", "name", "last.name",  
                  "age", "district", "moliner",  
                  "faculty", "year", "program",  
                  "prog.lang", "os", "labs",  
                  "Excel", "R", "why")
```

```
names(survey)
```

```
## [1] "date"      "name"      "last.name" "age"      "di  
## [7] "faculty"   "year"      "program"   "prog.lang" "os  
## [13] "Excel"    "R"         "why"
```



# De Excel a R: Vamos a leer la encuesta

## ► Cuántos estudian Ing. Ambiental

```
survey$faculty == "Ingeniería ambiental"
```

```
## [1] FALSE TRUE FALSE FALSE TRUE FALSE TRUE FALSE TH  
## [13] FALSE TRUE FALSE FALSE FALSE FALSE TRUE TRUE FAI
```

```
sum(survey$faculty == "Ingeniería ambiental")
```

```
## [1] 9
```

## De Excel a R: Vamos a leer la encuesta

- ▶ Cuántos estudian Ing. Ambiental, quienes son?

```
survey[survey$faculty == "Ingeniería ambiental", ]
```

##		date		name	last
## 2	3/25/2021 1:32:47			Meryann	LI
## 5	3/25/2021 1:52:05	Willian Francisco	Hidalgo de Cast		
## 7	3/25/2021 10:32:09			Evita	Me
## 9	3/25/2021 12:54:09			Pedro	Ca
## 11	3/25/2021 18:02:34			Pablo	
## 14	3/25/2021 19:48:54			Pedro	I
## 19	3/26/2021 12:10:41	Katherine Adriana	Zapana Ca		
## 20	3/26/2021 18:04:44	Melissa Olenka	Vilchez Fig		
## 24	3/27/2021 0:17:55			Alejandro	De
##		district		molinero	
## 2	Fuera de Lima	La Gris		No Ingeniería	
## 5	San Juan de Miraflores			No Ingeniería	
## 7	San Martín de Porres	Sí, y ya me gradué	Ingeniería		
## 9	Barranco		No Ingeniería		

## De Excel a R: Vamos a leer la encuesta

- ▶ Cuántos son menores que la edad média?

```
edad_clase <- mean(survey$age)
genz <- survey[survey$age < edad_clase, ]
nrow(genz)
```

```
## [1] 12
```

- ▶ O también

```
sum(survey$age < edad_clase)
```

```
## [1] 12
```

# De Excel a R: Vamos a leer la encuesta

- Qué estudia el más valiente?

```
names(survey)
```

```
## [1] "date"      "name"      "last.name" "age"      "di"
## [7] "faculty"   "year"      "program"   "prog.lang" "os"
## [13] "Excel"     "R"         "why"
```

```
par_fin <- "No hago las prácticas me defiendo en el parcial"
survey[survey$lab == par_fin, 7]
```

```
## [1] "Meteorología"
```

```
# Mejor sintaxis
```

```
survey$faculty[survey$lab == par_fin]
```

```
## [1] "Meteorología"
```

# De Excel a R: Vamos a leer la encuesta

- Qué estudia el más joven?

```
names(survey)
```

```
## [1] "date"      "name"      "last.name" "age"      "di"      "dis"
## [7] "faculty"   "year"      "program"   "prog.lang" "os"
## [13] "Excel"     "R"         "why"
```

```
survey[survey$age == min(survey$age), 7]
```

```
## [1] "Meteorología"
```

```
# Mejor sintaxis
```

```
survey$faculty[survey$age == min(survey$age)]
```

```
## [1] "Meteorología"
```

## De Excel a R: Vamos a leer la encuesta

- ▶ Cuántos no hicieron la tarea?

```
num_alumnos_total <- 28  
num_alumnos_enc <- nrow(survey)  
  
num_alumnos_total - num_alumnos_enc  
  
## [1] 4
```

# Hora de hacer el Hand-on lab 1!

- ▶ **Algoritmo de Feynmann** para la solución de problemas
  - ▶ Escribe el **problema**
  - ▶ **Piensa** mucho
  - ▶ Escribe la **solución**