

Hello Bioinformatics, in R

Taller de Bioinformática nivel introductorio

CodeMyGen by GenoBit

Eduardo Cepeda, Angel Peña, Sofia Acuña.

CodeMyGen

Talleres ofrecidos por **GenoBit: Genomics and Bioinformatics**
Para aquellos que buscan desarrollar habilidades de programación,
aplicada a diferentes ámbitos.

CodeMyGen

Talleres ofrecidos por **GenoBit: Genomics and Bioinformatics**
Para aquellos que buscan desarrollar habilidades de programación,
aplicada a diferentes ámbitos.

En la edición de hoy tenemos colaboración

CodeMyGen

Talleres ofrecidos por **GenoBit: Genomics and Bioinformatics**
Para aquellos que buscan desarrollar habilidades de programación,
aplicada a diferentes ámbitos.

En la edición de hoy tenemos colaboración
GenoBit x SEIBT

CodeMyGen

Ruta de aprendizaje

Introductorio



Hello, Bioinformatics
in R

Intermedio



Stay tuned...

Avanzado



Stay tuned...

Overview

- > R y RStudio
- > Manipulación de datos
- > Inspección y limpieza
- > Visualización
- > Case Analysis 

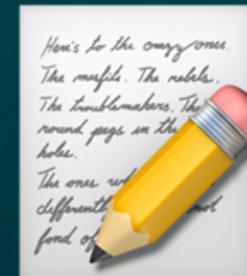
Overview

- > R y RStudio
- > Manipulación de datos
- > Inspección y limpieza
- > Visualización
- > Case Analysis 

Librerías

- > tidyverse
- > ggplot
- > pheatmap

R y RStudio, ¿Es lo mismo?



RStudio

R y RStudio, ¿Es lo mismo?



R

Es el lenguaje de programación, quien permite hacer los cálculos y comandos

RStudio

IDE con interfaz gráfica agradable visualmente, no es quien ejecuta código, solo lo hace verse bien

Descarga R y RStudio



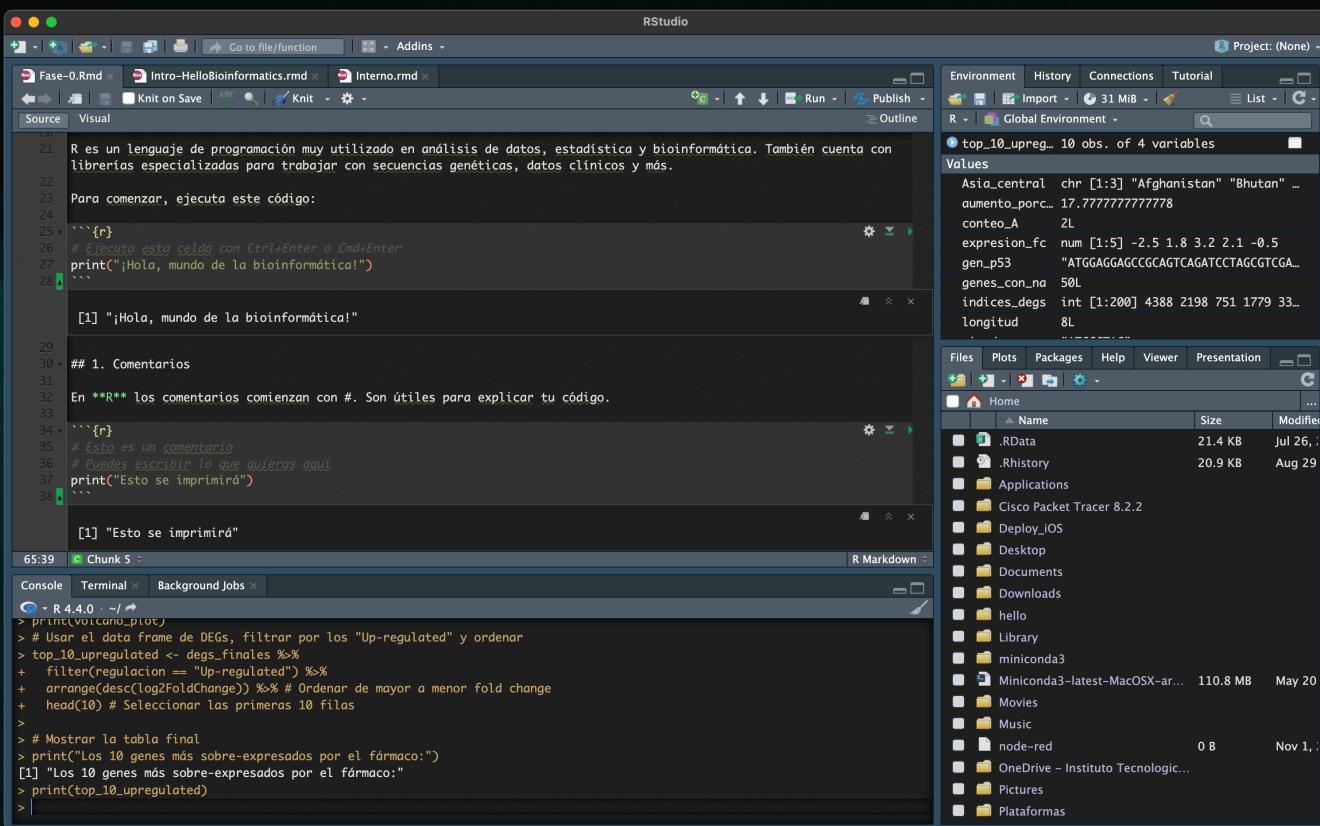
- > Descarga R
- > Descarga RStudio para Windows o Mac

Carpeta de Drive con cuadernillos



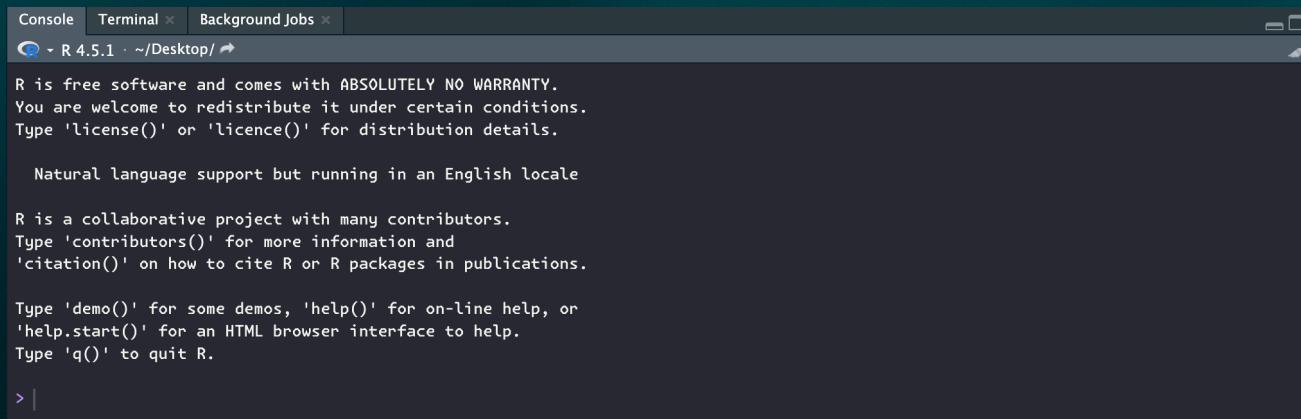
- > Descarga el archivo en el que trabajaremos
- > HelloBioinformatics.rmd

Layout de RStudio



Instalar Librerias

```
> install.packages("tidyverse")
> install.packages("pheatmap")
```



A screenshot of an R console window. The title bar shows "Console", "Terminal", and "Background Jobs". The main area displays the standard R startup message:

```
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.  
  
Natural language support but running in an English locale  
  
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.
```

The bottom of the window shows a command prompt with a greater than sign (>) followed by a vertical line.

Tipos de Datos

Numeric	Números decimales	numero <- 1.0
Integer	Números enteros	int <- 1
Character	Cadenas de texto	str <- "un texto"
Complex	Números complejos	comp <- 3+2i
Logical	Verdadero (TRUE) o falso (FALSE). Es a menudo el resultado de operaciones lógicas.	a <- 1; b <- 2; a < b

https://bookdown.org/chescosalgado/intro_r/tipos-y-estructuras-de-datos-en-r.html

Elemento/Variable

```
object_name <- value
```

<https://r4ds.had.co.nz/workflow-basics.html>

Vectores

```
x <- c("Austria", "Tyrol", "Innsbruck")
y <- c("Germany", "Berlin", "Berlin")
z <- c(x, y)
z

## [1] "Austria"    "Tyrol"       "Innsbruck"   "Germany"    "Berlin"     "Berlin"
```

```
seq(1, 10)
#> [1] 1 2 3 4 5 6 7 8 9 10
```

<https://r4ds.had.co.nz/workflow-basics.html>

<https://discdown.org/rprogramming/vectors.html#vectors-creating-vectors>

Data Frame

```
data <- data.frame(month = month, temperature = temp,  
                    humidity = humidity, rain = rain)
```

	month	temperature	humidity	rain
1	January	20.37	88	72.0
2	February	18.56	86	33.9
3	March	18.40	81	37.5
4	April	21.96	79	36.6
5	May	29.53	80	31.0
6	June	28.16	78	16.6

<https://r-coder.com/dataframe-r/>

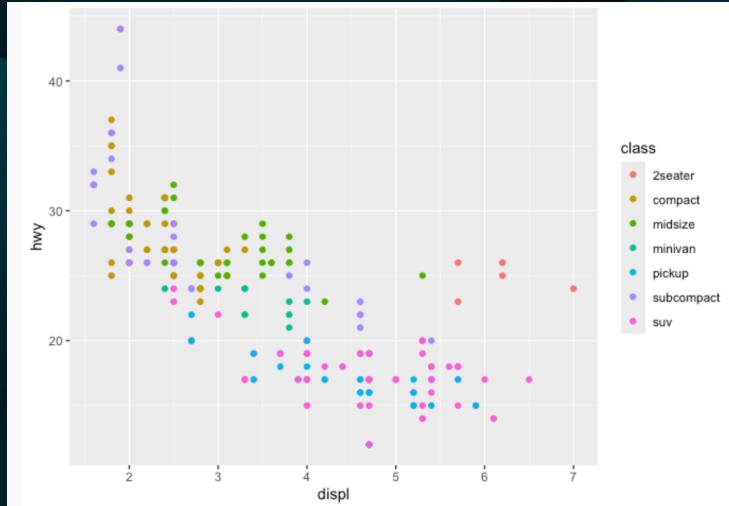
Inspección y Limpieza (tidyverse)

```
filter(flights, month == 1, day == 1)
#> # A tibble: 842 × 19
#>   year month   day dep_time sched_dep_time dep_delay arr_time sched_a
#>   <int> <int> <int>     <int>          <int>     <dbl>    <int>
#> 1  2013     1     1      517            515       2     830
#> 2  2013     1     1      533            529       4     850
#> 3  2013     1     1      542            540       2     923
#> 4  2013     1     1      544            545      -1    1004
#> 5  2013     1     1      554            600      -6     812
#> 6  2013     1     1      554            558      -4     740
#> # i 836 more rows
#> # i 11 more variables: arr_delay <dbl>, carrier <chr>, flight <int>,
#> #   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>, distance
#> #   hour <dbl>, minute <dbl>, time_hour <dttm>
```

<https://r4ds.had.co.nz/transform.html>

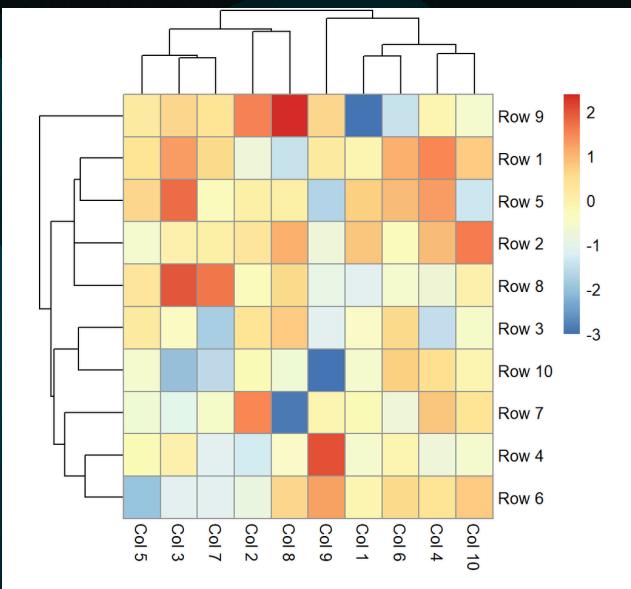
Visualizar los datos (ggplot)

```
ggplot(mpg, aes(displ, hwy, colour = class)) +  
  geom_point()
```



<https://ggplot2.tidyverse.org/>

HeatMap



```
# Data
set.seed(8)
m <- matrix(rnorm(200), 10, 10)
colnames(m) <- paste("Col", 1:10)
rownames(m) <- paste("Row", 1:10)

# Heat map
pheatmap(m)
```

<https://r-charts.com/correlation/pheatmap/>

Case Analysis

Contexto: Somos bioinformáticos analizando un experimento de **RNA-seq**. Hemos comparado la expresión génica de células cancerosas tratadas con un nuevo fármaco contra células no tratadas (control). Tenemos un archivo con los resultados: el nombre del gen, el cambio de expresión promedio (`log2FoldChange`) y un p-valor ajustado que indica la significancia de ese cambio.

Objetivo: Identificar y visualizar los genes diferencialmente expresados (DEGs) para entender el efecto del fármaco. Un gen se considera DEG si $\text{log2FoldChange} > 1$ (o < -1) y $p_{\text{adj}} < 0.05$.

Feedback

