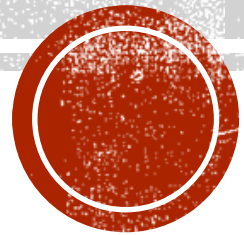
INTRODUCCIÓN A R

Christian Camilo Urcuqui López, MSc

Github: urcuqui
https://github.com/urcuqui



UNIVERSIDAD ICESI | A OTRO NIVEL

PRESENTACIÓN

Christian Camilo Urcuqui López

Ing. Sistemas, Magister en Informática y Telecomunicaciones

Big Data Professional

Big Data Scientist

Deep Learning Specialization

Grupo de investigación i2t

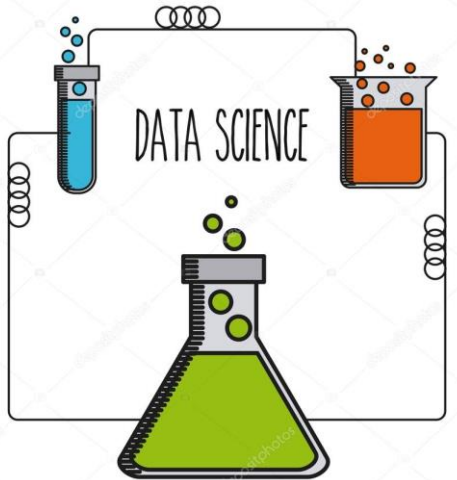
Líder de investigación y desarrollo

Ciberseguridad y ciencia de datos aplicada

ccurcuqui@icesi.edu.co

COMPETENCIAS

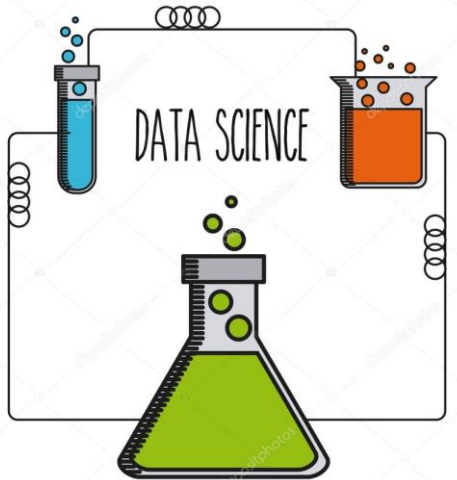
- Describir el lenguaje de programación R y su aplicación en proyectos de ciencia de datos
- Aplicar los conceptos básicos de codificación en R
- Explorar un *data.frame* en R



UNIVERSIDAD ICESI | A OTRO NIVEL

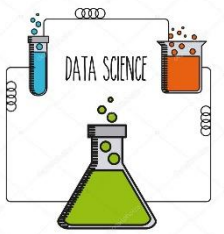
CONTENIDO

1. Introducción al lenguaje de programación R
2. Flujo de trabajo en R - Nombres de objetos, llamada a funciones
3. Explorando las estructuras de datos en R
4. Importando y trabajando con datos
5. Resolución de dudas

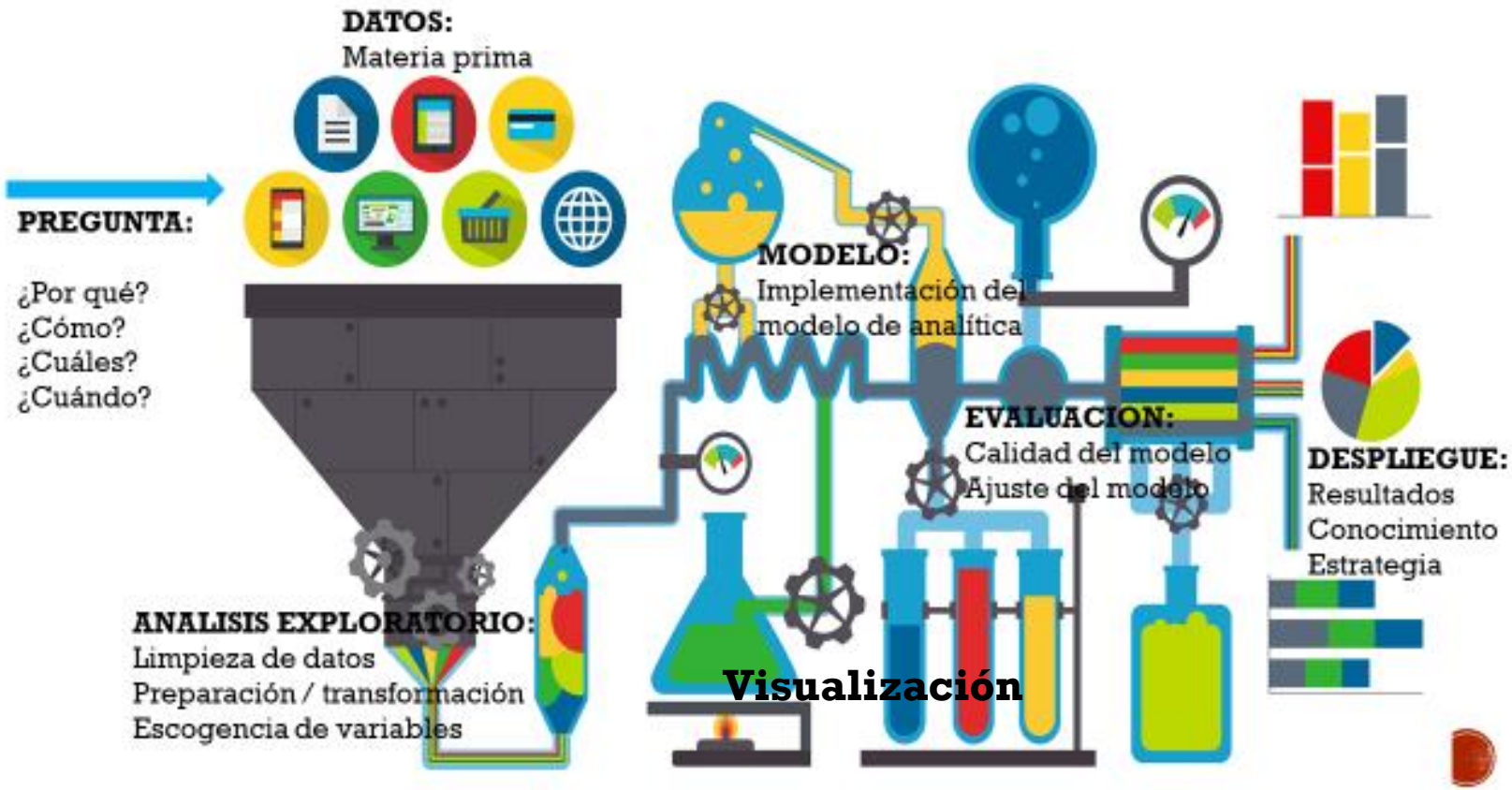


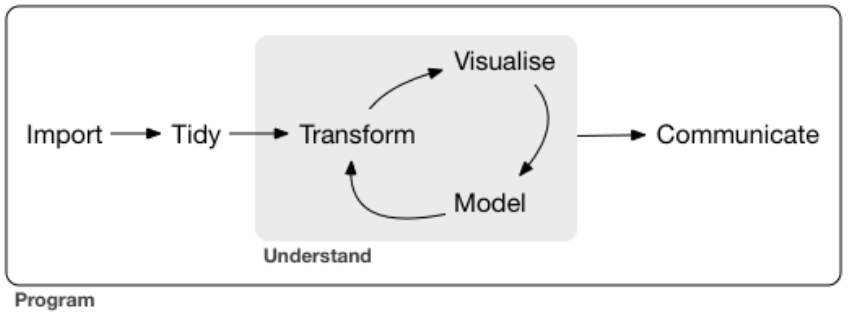
INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN R

- La forma de analizar los datos ha estado cambiando durante los últimos años.
- "Los datos son el petróleo del siglo xxi".
- La ciencia de datos a partir de sus técnicas (por ejemplo, estadísticas, visuales, econométricas y de aprendizaje de máquina) han permitido descubrir y explotar la información.
- Antiguamente, los investigadores solían publicar sus resultados en prestigiosas revistas y la implementación de sus descubrimientos en software tomaba mucho tiempo.
- Actualmente, los investigadores y la industria han mejorado sus métodos en conjunto con la implementación de software, estos resultados ahora se encuentran en sitios web de fácil acceso (en muchos casos con licencia *open source*).

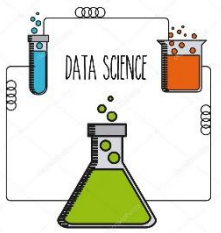


RECORDEMOS

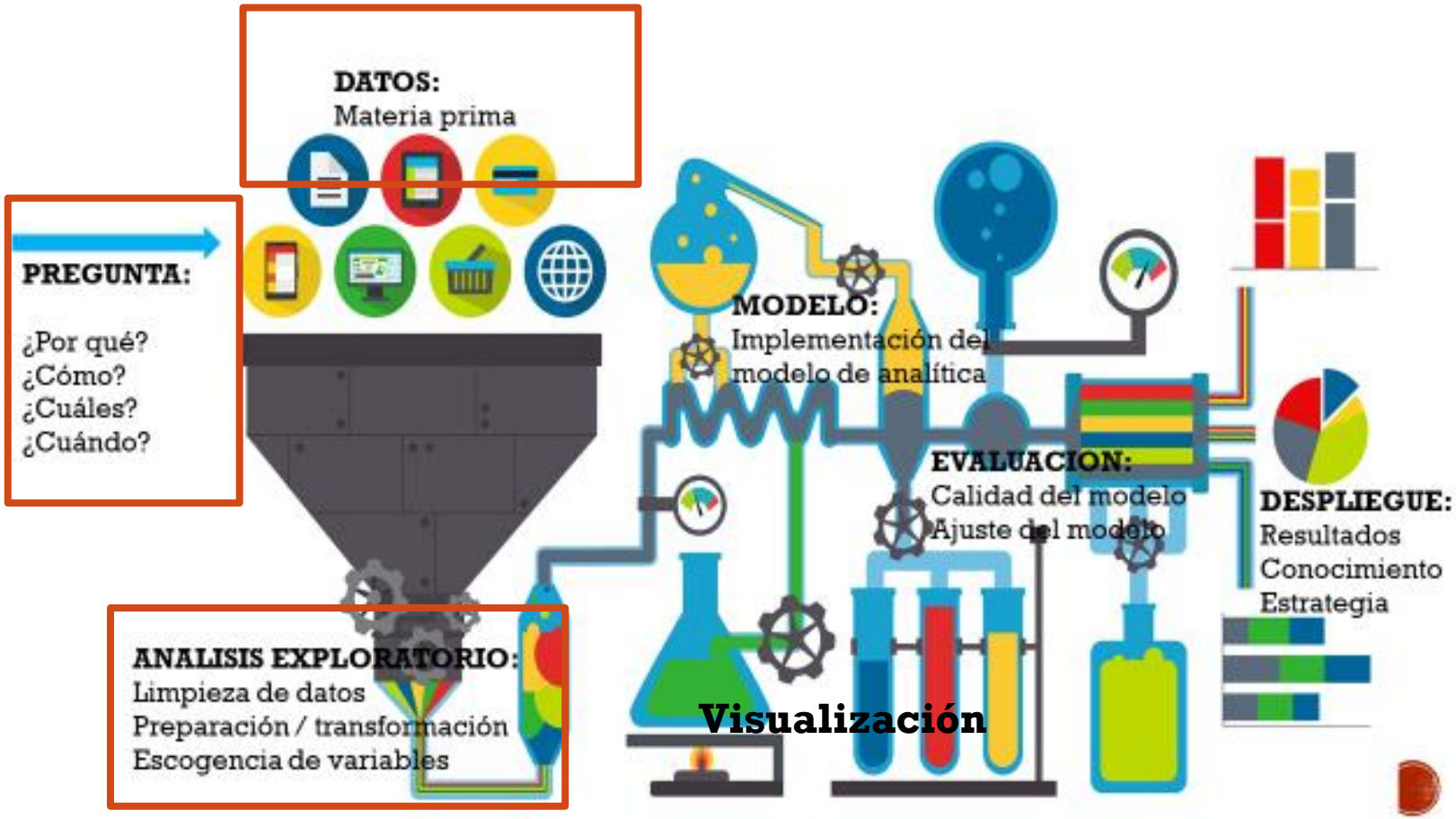


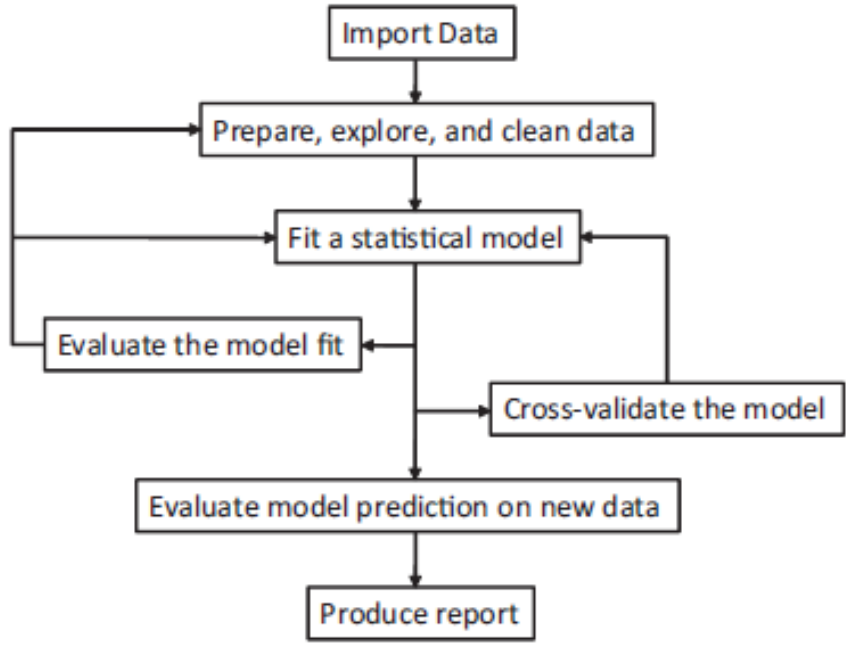


Marco de trabajo típico de un proyecto de ciencia de datos.
R for Data Science



RECORDEMOS



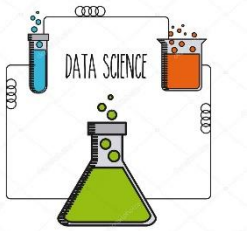


Pasos típicos de un proyecto de ciencia de datos.

R IN ACTION: Data analysis and graphics with R

INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN R

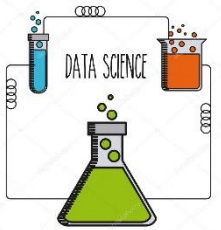
- R es un entorno y un lenguaje para computación estadística y gráfica desarrollado en Bell Labs, un proyecto que nace del software libre S.
- Una solución *open source* para análisis de datos soportado por varias comunidades científicas en todo el mundo.
- ¿Por qué utilizar R en vez de otras soluciones populares para estadística y gráficos (por ejemplo, Microsoft Excel, SAS, IBM SPSS, Stata, y Minitab)?



INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN R

Excel

- **Cálculo básicos.** Excel provee una interfaz (entorno) más amigable para cálculos simples (por ejemplo, estadísticas descriptivas) o algunas manipulaciones sencillas (por ejemplo, filtros y búsquedas).
- **Ver los datos continuamente**. Excel es una herramienta que nos permite constantemente ver la estructura y el contenido de los datos.
- **Presentación de datos y resumen**. Excel nos da un contenido estético más agradable de las hojas de cálculo.
- **Menor curva de aprendizaje**. Requiere un menos tiempo para llegar a manejar gran parte de sus funcionalidades. n



INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN R

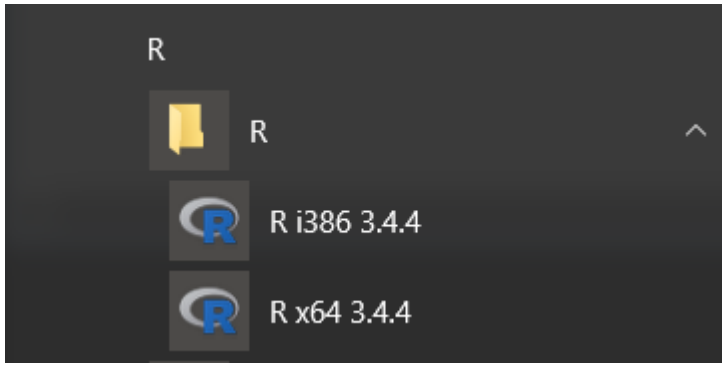
R

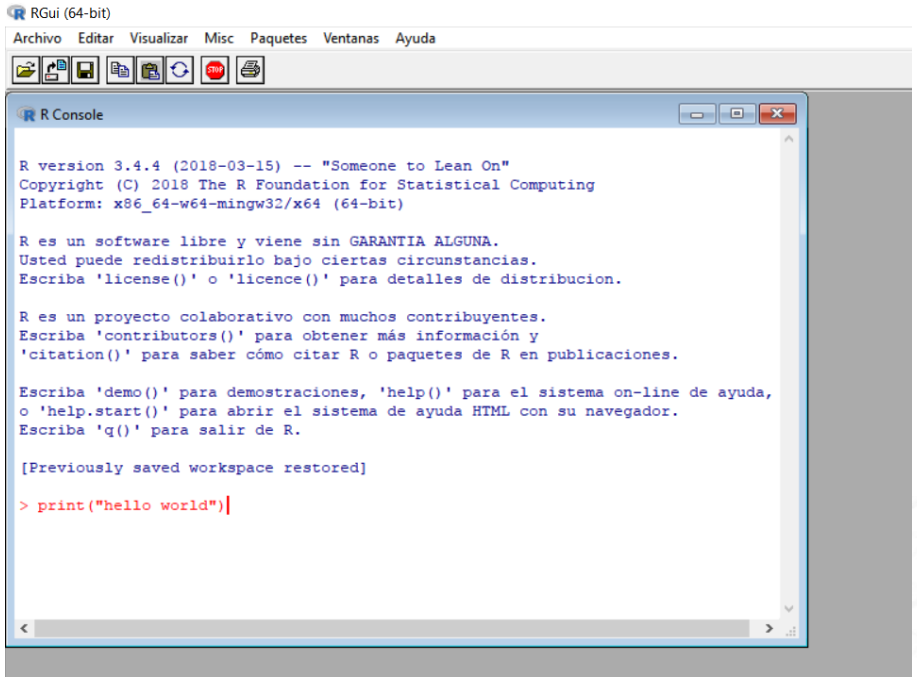
- Es una tecnología gratuita a diferencia de otros software comerciales de altos costos.
- Tiene comunidades muy activas, casi semanalmente se proponen nuevos paquetes estadísticos y actualizaciones, lo cual a llevado a los sistemas comerciales a integrar R.
- Es más fácil la automatización ya que se pueden desarrollar *scripts (líneas de código con un propósito especifico)* que permiten ejecutar el análisis varias veces.
- Leer casi cualquier tipo de datos (.txt, .csv, .dat), también, existen paquetes que permiten leer información de archivos JSON, Excel, STATA, SAS. E incluso utilizar datos de sitios web y de sistemas de base de datos (Por ejemplo, MySQL y, PostgresSQL)

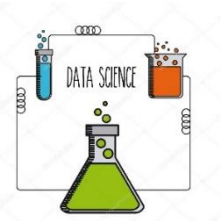
INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN R

Nuestro primer hello world

- Una vez instalado R en nuestro equipo, nosotros podemos acceder a la consola de R para que podamos escribir los código. Procedamos ha abrir una consola y digitemos el comando **print("hello world")**





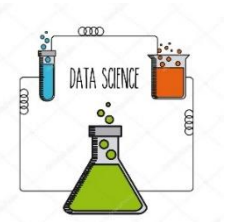


INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN R

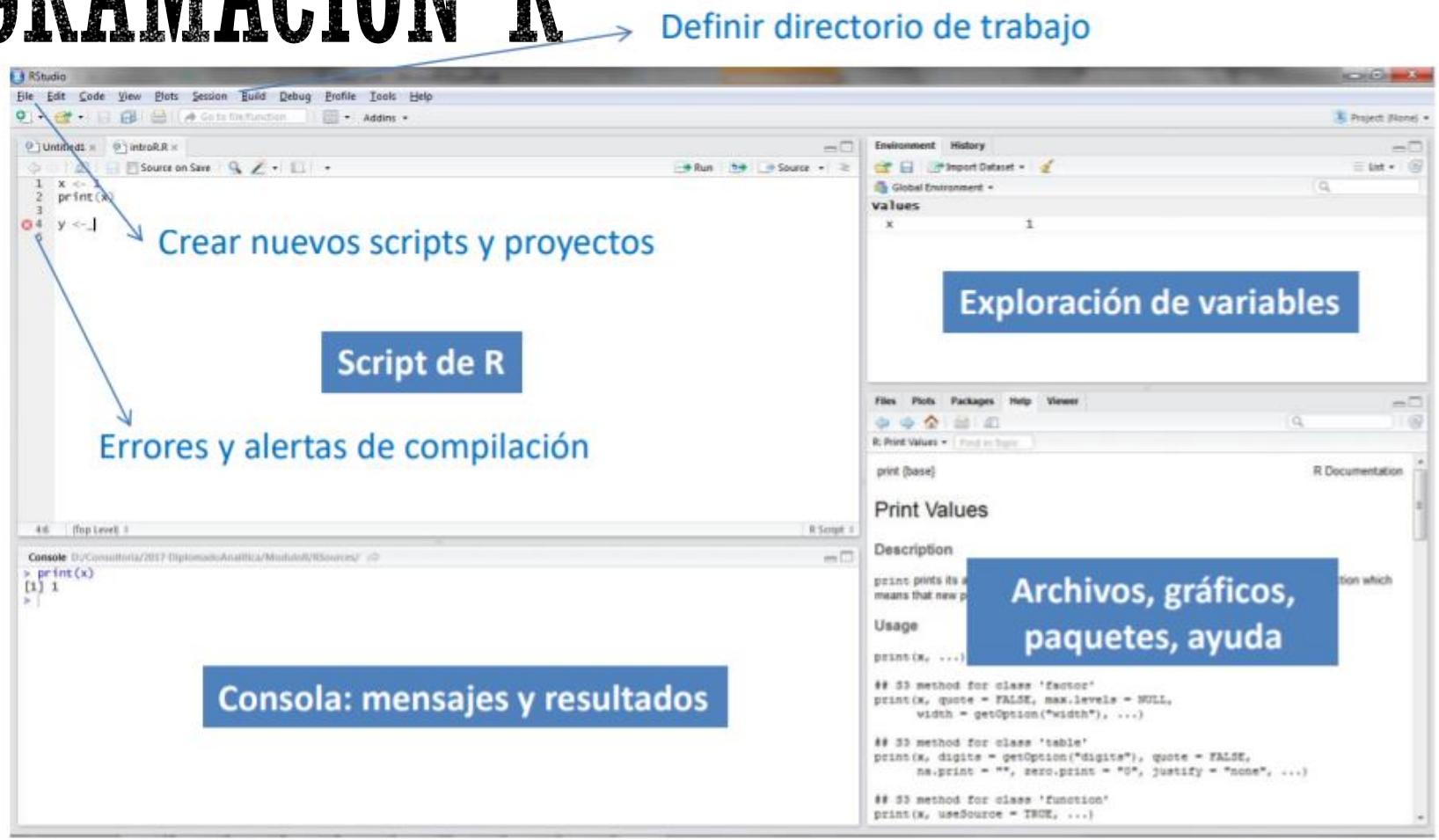
- Existen ambientes de programación más amigables y con una serie de utilidades que nos facilitan el desarrollo de proyectos de software, para nuestro caso utilizaremos RStudio.
- RStudio puede descargarse de la siguiente página web:

https://www.rstudio.com/products/rstudio/download/

- RStudio nos facilita el trabajo con R
  - Editor de código
  - Depurador (permite probar y depurar errores en tiempo de ejecución)
  - Herramientas de visualización
- Algunas de sus versiones son de uso libre, otras son licenciadas.

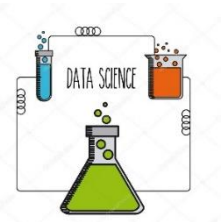


INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN R



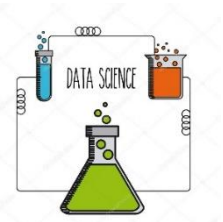
INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN R

- R es un lenguaje de programación interpretado, es decir, el código será ejecutado instrucción por instrucción.
- Muchos de los datos y variables son almacenadas en memoria durante una sesión. Nosotros podemos guardar una sesión con la finalidad de conservar nuestro trabajo para futuras sesiones.
- R utiliza el símbolo **<-** para detonar una asignación, a diferencia del típico = utilizado en muchos otros lenguajes de programación.
- ¡Veamos el ejemplo en el notebook!



INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN R

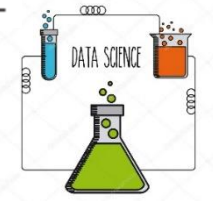
- Como hemos mencionado, la aplicación de la función rnorm nos dio un vector de cinco valores, ahora si deseamos crear un vector y asignarlo a una variable, debemos proceder a digitar la función _c_ seguido de los valores que queremos tener.
- ¡Veamos el ejemplo en el notebook!
- Ahora, si deseamos obtener mayor información sobre una función podemos utilizar la opción de ayuda con R a través de **?**,
- ¡Veamos el ejemplo en el notebook!



INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN R

- Los comentarios en el código podemos agregarlos con #
- Procedamos a un tercer ejercicio, vamos a crear una tabla a través de la asignación de vectores, utilizaremos unas funciones predefinidas en R para estadística descriptiva y vamos a crear un gráfico (conocidos como -*plot*).

Age (mo.)	Weight (kg.)	Age (mo.)	Weight (kg.)
01	4.4	09	7.3
03	5.3	03	6.0
05	7.2	09	10.4
02	5.2	12	10.2
11	8.5	03	6.1



INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN R

- Existen numerosas funciones de R en distintos paquetes que nos pueden facilitar la vida. La instalación de un paquete se realiza a través de:

  install.packages("nombre del paquete")

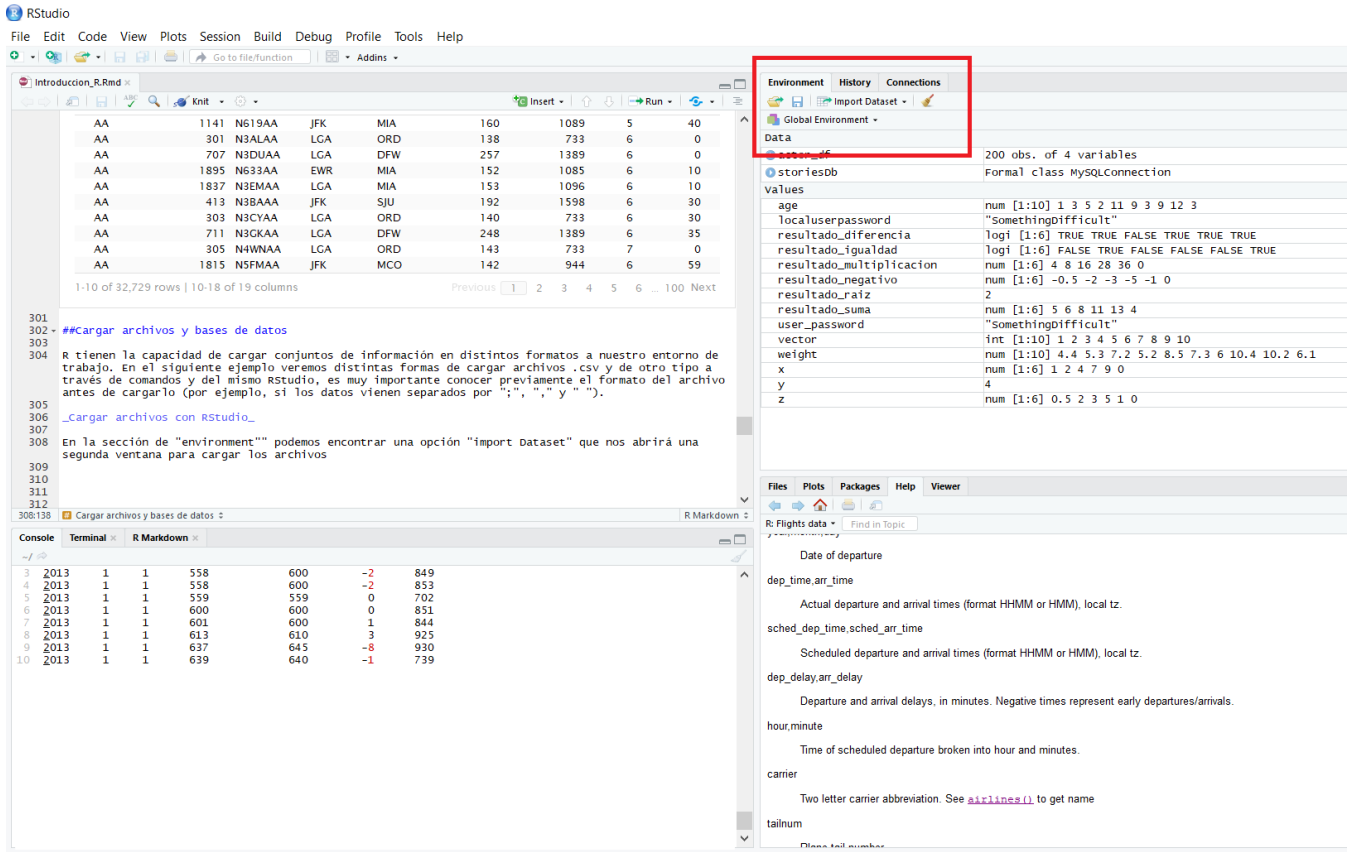
- Como hemos mencionado, constantemente las comunidades lanzan nuevas versiones de sus paquetes ya sea para incorporar nuevas funcionalidades o para corregir algún error; si deseamos estar en la ultima versión podemos utilizar el siguiente comando:

  update.packages()

- Para cargar el paquete que necesitamos utilizar en nuestro entorno de trabajo debe utilizar el siguiente comando en conjunto con el nombre del paquete

  library(gclus)

INTRODUCCIÓN AL LENGUAJE DE PROGRAMACIÓN R



REFERENCIAS

- Kabacoff, R. (2015). R IN ACTION: Data analysis and graphics with R.
- Wickham, H., & Grolemund, G. (2016). R for Data Science.