

¡ BIENVENIDXS !

CLUB DE LECTURA

FUNDAMENTOS DE CIENCIA DE DATOS CON R

Capítulos: 1,2,3



Ladies



@rladiesbquilla

Ladies

AGENDA



- ✓ ¿Qué es R-Ladies?
- ✓ Nuestras expertas de hoy.
- ✓ Presentación libro **Fundamentos de ciencia de datos con R**.
- ✓ ¿Qué es la ciencia de datos? ¿Hay qué aplicar alguna metodología?
- ✓ Introducción a R y RStudio.
- ✓ Estructuras y tipos de datos.
- ✓ Coerción de tipos.





Countries
63

Chapters
219

Events
3943



@rladiesbquilla



R-Ladies Barranquilla



barranquilla@rladies.org



https://github.com/r-ladies/mee_tup-presentations_barranquilla



CÓDIGO DE CONDUCTA



Brindamos una experiencia libre de acoso para todos. No toleramos el acoso de los participantes de ninguna forma.



Evita realizar comentarios ofensivos relacionados con el género, la identidad y expresión de género, la orientación sexual, la discapacidad, la enfermedad mental, la neuro(a)tipicidad, la apariencia física, el tamaño corporal, la edad, la raza o la religión.



Si está siendo acosadx por un miembro/invitadx/participante de R-Ladies, nota que alguien más está siendo acosadx o tiene alguna otra inquietud, comuníquese inmediatamente con nuestro equipo.



<https://rladies.org/coc/>

NUESTRAS EXPERTAS DE HOY



Isabel Vasquez
Analista de datos



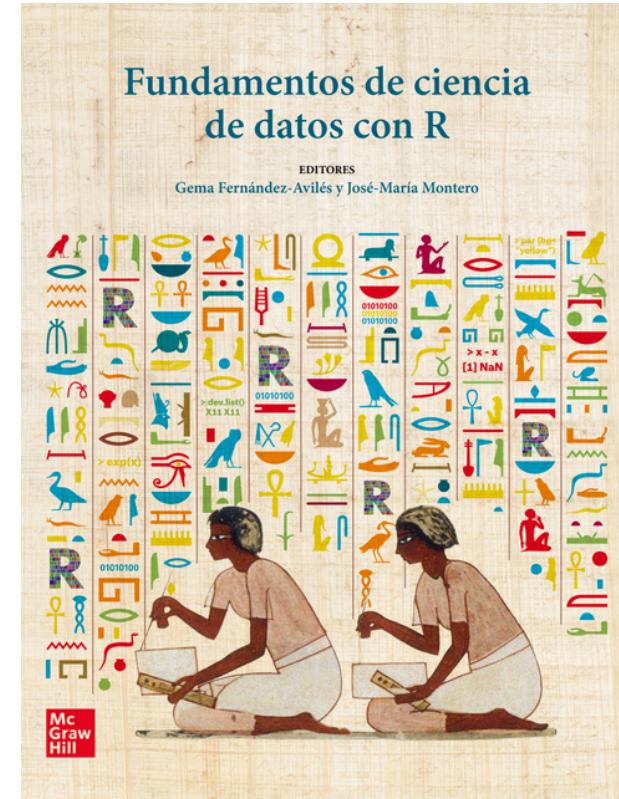
Viviana Flórez
Científica de datos

FUNDAMENTOS DE CIENCIA DE DATOS CON R

- ✓ Completitud
- ✓ Amplitud técnica
- ✓ Reproducibilidad
- ✓ Variedad de perspectivas



DESEO DE RESOLVER PROBLEMAS
UTILIZANDO DATOS



¿QUÉ ES R?

ES UN LENGUAJE DE CÓDIGO ABIERTO PARA COMPUTACIÓN ESTADÍSTICA. LA UTILIDAD DE R PARA LA CIENCIA DE DATOS DERIVA DE UN FANTÁSTICO ECOSISTEMA DE PAQUETES (ACTIVO Y EN CRECIMIENTO)



¿QUÉ ES LA CIENCIA DE DATOS?



Implica la limpieza, la agregación y la manipulación de datos de diversas fuentes

La Ciencia de Datos es uno de los campos de más rápido desarrollo en todas las industrias.

Combina la estadística, las matemáticas, la informática y la programación, para obtener valor de los datos.

METODOLOGÍA EN CIENCIA DE DATOS

KDD

Sistematizar el proceso conocido hoy día como ciencia de datos.

SEMMA

Proceso de selección, exploración y modelización de grandes bases de datos.

CRISP-DM

Seis fases que permiten la construcción e implementación de un modelo de minería de datos

CRISP-DM

6.
Implementación.

5.
Evaluación.

4.
Modelado.

1.
Entendimiento del
negocio.

2.
Comprensión de los
datos.

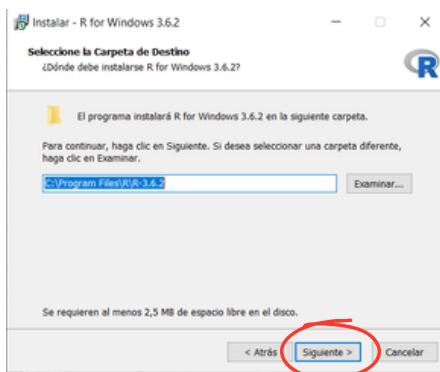
3.
Preparación de los
datos.

INSTALAR R

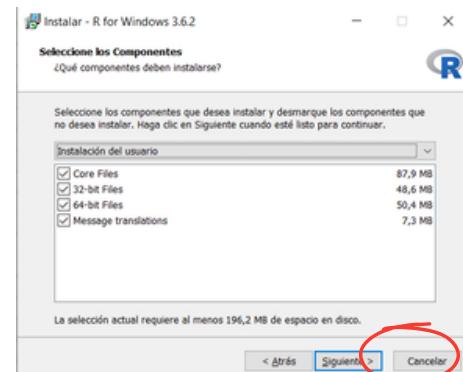
1. <https://cran.r-project.org/bin/windows/base/>



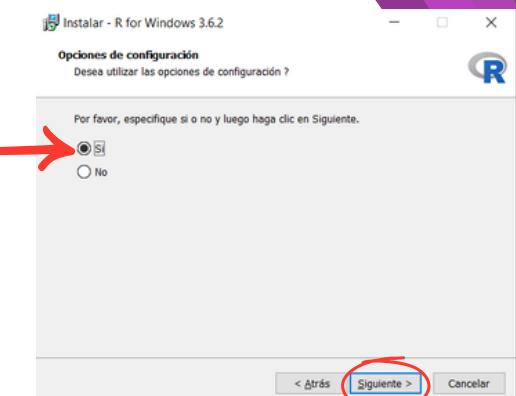
2.



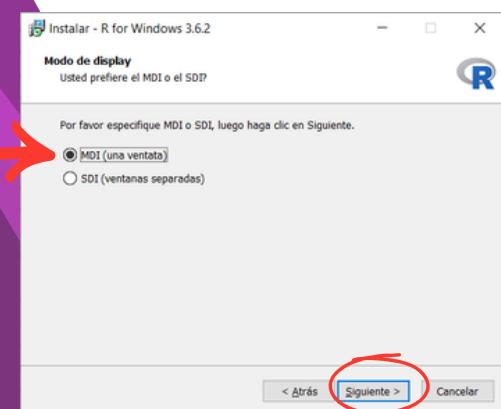
3.



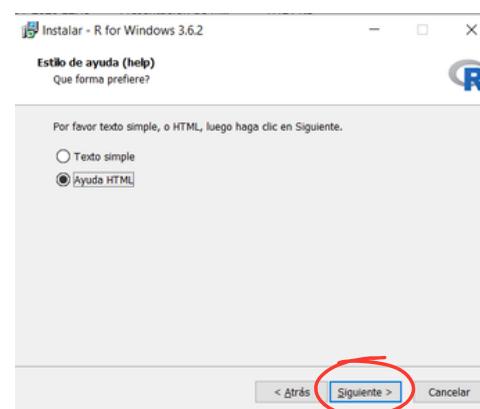
4.



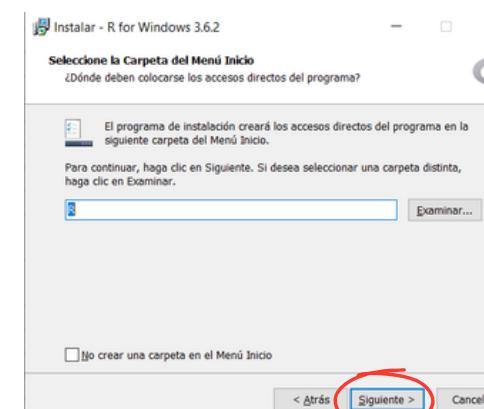
5.



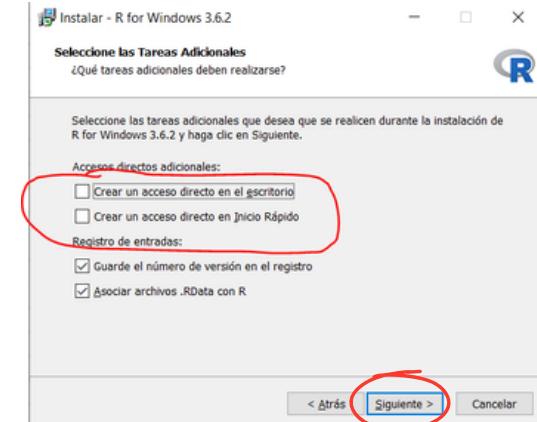
6.



7.

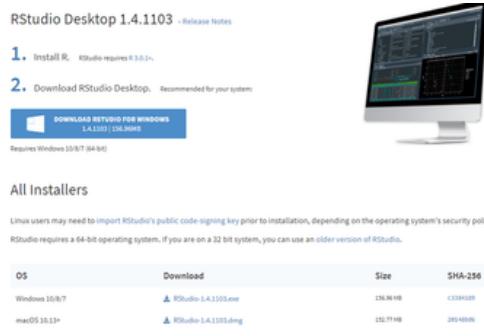


8.

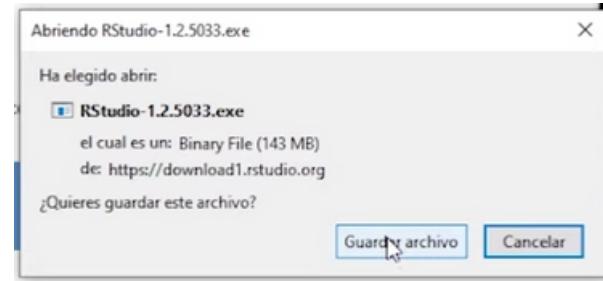


INSTALAR R Studio®

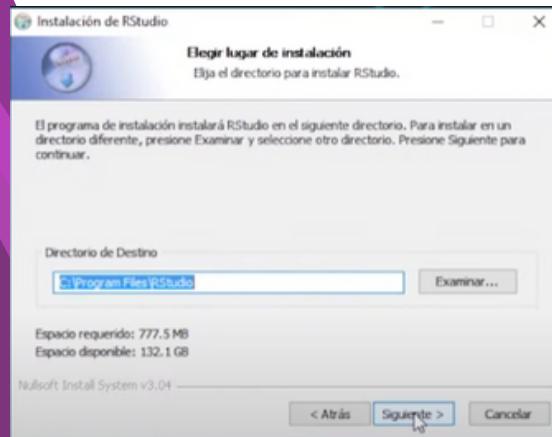
1. [https://posit.co
/download/rstu
dio-desktop/](https://posit.co/download/rstudio-desktop/)



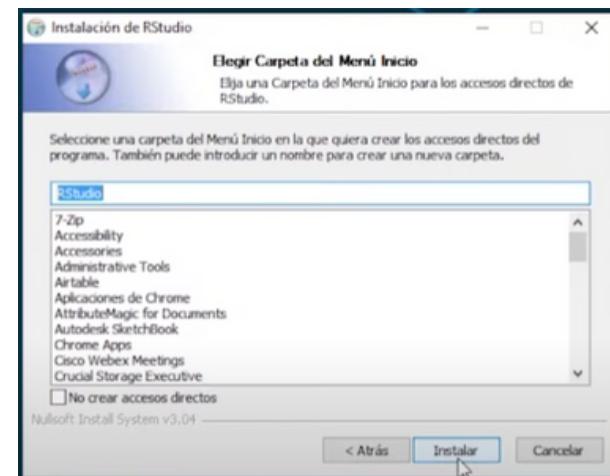
2.



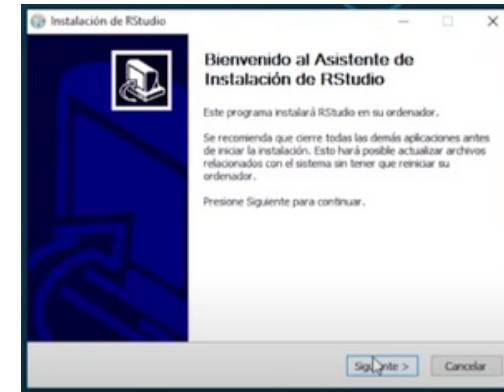
4.



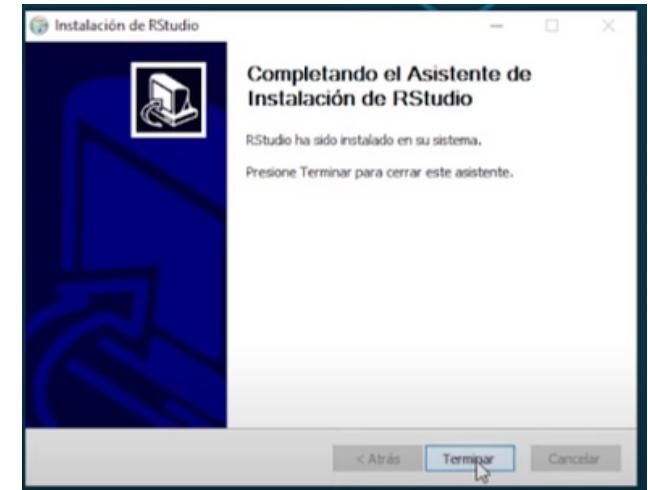
5.



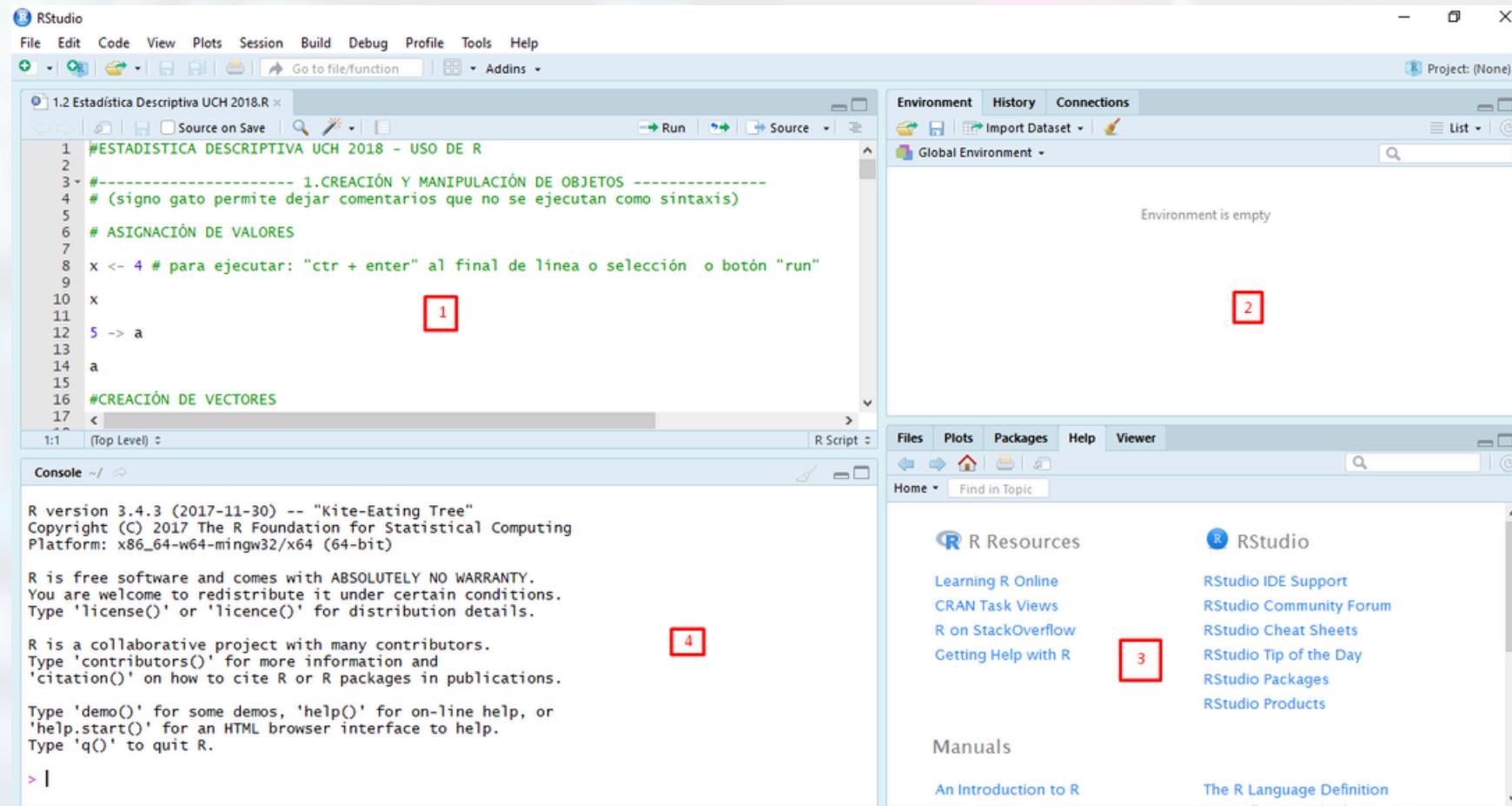
3.



6.



ENTORNO DE R Studio



CÓDIGO LIMPIO

COMENTARIOS



ESCRIBIR BUENOS COMENTARIOS ES MÁS UN ARTE QUE UNA CIENCIA.

CÓDIGO COMPARTIDO

INCLUSO SI TRABAJA SOLO, PROBABLEMENTE LE GUSTARÍA QUE SU PROGRAMA SE EJECUTE CORRECTAMENTE EN OTRAS MÁQUINAS.



VARIABLES



NOMBRES DE VARIABLES ESPECÍFICOS

BUCLLES

COMPROBAR SIEMPRE SI EL BUCLE ES UNA BUENA OPCIÓN.



Símbolo	Uso	Paquete	Descripción detallada	Uso en manipulación de datos
%>% . >	Pipe	dplyr	Pasa el resultado de una expresión como argumento a la siguiente expresión.	Encadena múltiples operaciones de manipulación de datos, facilitando la lectura y estructura del código.
+, -, *, /	Operadores aritméticos	Base R	Realizan operaciones aritméticas básicas como suma, resta, multiplicación y división.	Se utilizan para realizar cálculos entre columnas o crear nuevas columnas calculadas en un dataset.
==, !=	Operadores de igualdad y diferencia	Base R	Comparan valores para determinar si son iguales o diferentes.	Se emplean en filtros o condicionales para seleccionar o excluir datos basados en condiciones específicas.
<, <=, >, >=	Operadores de comparación	Base R	Comparan valores para determinar relaciones de menor que, mayor que, etc.	Utilizados en la filtración y ordenación de datos para comparar valores numéricos o de fecha.
&,	Operadores lógicos AND - OR	Base R	Realizan operaciones lógicas AND y OR entre vectores de manera elemento a elemento.	Combinan condiciones en filtros para seleccionar datos que cumplen con múltiples criterios.
%in%	Operador de igualdad	Base R	Verifica si los elementos de un vector se encuentran en otro vector.	Se utiliza para filtrar filas cuyos valores de columna están dentro de un conjunto específico de valores.
:	Sequence operator	Base R	Genera una secuencia de números enteros.	Se utiliza para crear rangos de índices cuando se subconjuntan o se seleccionan filas específicas.

CLUB DE LECTURA

online

FUNDAMENTOS DE CIENCIA DE DATOS CON R

Capítulo: 3.5.2 - 3.6.2.3

 Tratamiento de datos con R

 Transformación de datos con dplyr

 Importación y exportación de datos

 Operaciones de resumen y agrupación



22 JUNIO, 2024

13:00 - 15:00 UTC-5

 RLadies Barranquilla

 zoom

Mary Rivero
Analista de datos

Danisse Carrascal
Ingeniera Industrial



¡IMAGAMOS ARTE!



ALLISON HORST