

Unidad 3. Manipulación de datos en R

Del Curso introductorio al lenguaje de programación R orientado al análisis cuantitativo en Ciencias Sociales por Sarahí Aguilar González

Objetivo: Que el estudiante sea capaz de importar y exportar bases de datos de y en diferentes formatos, así como ejecutar transformaciones básicas sobre estas utilizando data frames y el paquete `data.table`.

Duración: 2 sesiones (4 horas)

Agenda



Paquetes de R



Data Frames

Agenda



Paquetes de R



Data Frames

corre en



Lenguaje de programación y entorno
enfocado al cómputo estadístico y
graficación.

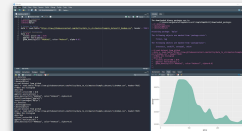
```
R >= 2009.09
df = data.frame(m=rep("m", 1000), y=rep(1, 1000),
               p=rep("p", 1000), s=rep("s", 1000),
               g=rep("g", 1000))
ggplot(df, aes(x, y)) +
  geom_density(aes(x, y)) +
  theme_minimal()
```

R Base

La colección de funciones R que se carga cada vez que inicia
R. Estas funciones proporcionan los conceptos básicos del
lenguaje y no es necesario cargar un paquete antes de poder
usarlas.



Entorno de desarrollo integrado (IDE) para R.



No eres la única persona que escribe sus propias funciones con R.

Muchos profesores, investigadores y programadores usan R para construir y actualizar sus propias funciones que se adecuan a sus necesidades.

Algunos, empaquetan sus funciones y las comparten con el mundo.

R promueve esto y lo permite hacer de forma fácil.



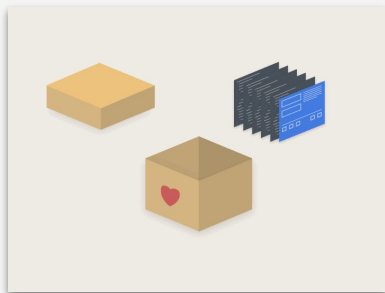
Un paquete R agrupa funciones, conjuntos de datos y archivos de ayuda.

Cualquiera puede usar estas **funciones** dentro de su propio código R una vez que cargue el paquete en el que residen.

Por lo general, el contenido de un paquete R **está relacionado con un solo tipo de tarea**, que el paquete ayuda a resolver.

Los paquetes necesitan ser **descargados** (una sola vez) y **cargados** al entorno de R (cada vez que se usarán).

Los paquetes son actualizados por sus creadores, y para correr correctamente en tu R, **deben ser compatibles con su versión**.



Paquetes de R

corre en



Lenguaje de programación y entorno
enfocado al cómputo estadístico y
graficación.

```
R >= 2009.02
df = data.frame(morechN(), morechN, 5, 2), morechN, 10, 33),
               p=morechN(), morechN, 5, 2), morechN, 10, 33),
               group=CC("R", "C"), each=4))

ggplot(df, aes(x, fill=gg)) +
  geom_density(aes(x=latitud, y=latitud)) +
  theme_minimal()
```

R Base

La colección de funciones R que se carga cada vez que inicia
R. Estas funciones proporcionan los conceptos básicos del
lenguaje y no es necesario cargar un paquete antes de poder
usarlas.

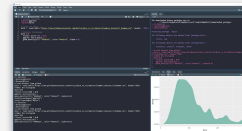
Paquetes de R

Una colección de funciones escritas en R descargables. Estas
funciones proporcionan conceptos más avanzada del
lenguaje de forma simplificada, pero es necesario
descargarlas y cargarlas antes de poder usarlas.

corre sobre



Entorno de desarrollo integrado (IDE) para R.



Paquetes de R

R Studio

tidyverse

dplyr

readr

ggplot2

tidyr

TIBBLE

¿Qué paquetes de R conocen o han utilizado?

knitr

rmarkdown

feather

purrr

lubridate

broom

Shiny

Instalando paquetes

(Descargar a tu computadora)

La forma más sencilla de instalar un paquete de R es con la función `install.packages("<package>")`

Esto buscará el paquete especificado en la colección de paquetes alojados en el sitio CRAN. Cuando R encuentre el paquete, lo descargará en una carpeta de bibliotecas en su computadora. R puede acceder al paquete aquí en futuras sesiones de R sin reinstalarlo.

Cargando paquetes

(Cargar sus funciones en tu sesión)

Para cargar un paquete de R, se utiliza la función R es con la función `library(<package>)`.

Esto pondrá a tu disposición todas las funciones, conjuntos de datos y archivos de ayuda del paquete hasta que cierre su sesión actual de R. La próxima vez que inicies una sesión de R, tendrás que volver a cargar el paquete si deseas usarlo, pero no tendrás que volver a instalarlo.

Buenas prácticas

- Utilizar Base R únicamente en medida de lo posible.
- Leer documentación.



Agenda

A vertical line runs down the left side of the slide. It has two small rectangular markers: a light gray one above a dark gray one.

Paquetes de R

Data Frames

Tipos de datos en R



Las 2 **propiedades** de un vector son:

Tipo



Tamaño



Además, pueden también contener **atributos adicionales** que los convierten en **vectores aumentados**.

Los **factores** están contruidos *encima* de los vectores numéricos de tipo entero.

Las **fechas** están contruidas *encima* de los vectores numéricos.

Los **data frames** están contruidos encima de los vectores de tipo lista.



Los data frames son la versión bidimensional de una lista y son la estructura de almacenamiento más útil para el análisis de datos.

Los data frames agrupan los vectores en una tabla bidimensional. Cada vector se convierte en una columna de la tabla.

Como resultado, cada columna de un data frame puede contener un tipo de datos diferente; pero dentro de una columna, cada celda debe tener el mismo tipo de datos.

1	"R"	TRUE
2	"S"	FALSE
3	"T"	TRUE