

1. Paqueterías/Paquetes
2. Importar datos
3. Manejo de datos
4. Introducción a tidyverse
5. Live coding

Paqueterías/Paquetes

- Los paquetes son un **conjunto de funciones**
- Algunos están integrados automáticamente en R.
 - base
 - utils
- Otros tienen que ser cargados manualmente y fueron creados por algún usuario para resolver un objetivo específico.
 - readr
 - tidyverse

- Para **instalar** un nuevo paquete escribe en la consola:

```
> install.packages("package.name")
```

Esto sólo es necesario una vez

- Para **cargar** un paquete:

```
> library(package.name)
```

Importar datos



Al iniciar una sesión nueva en R y crear un nuevo script:

1. Limpiar nuestro *environment*:

```
rm(list = ls())
```

2. Definir las paqueterías que queremos utilizar:

- Identificar tipo de datos que queremos leer
- Identificar si queremos calcular estadísticas, graficar etc.

Es buena práctica cargar las paqueterías al principio de código

3. Definir nuestro directorio de trabajo (Dónde leeremos datos/guardaremos resultados/código)

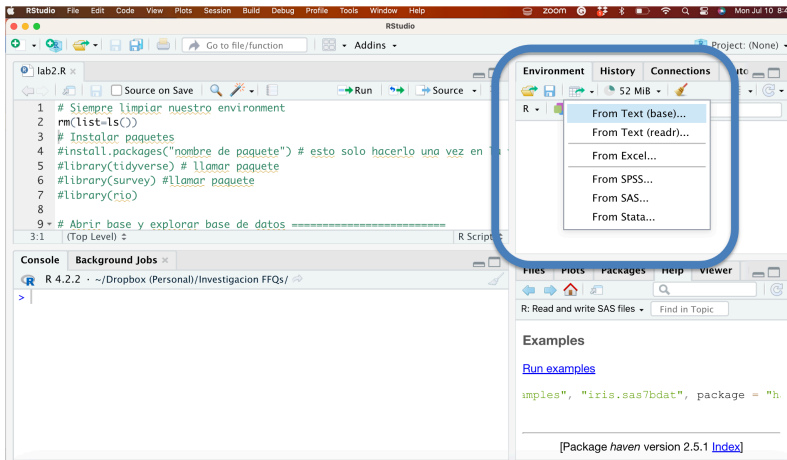
```
setwd("Carpeta")
```

4. Verificar nuestro directorio de trabajo

```
getwd()
```

Importar datos

Opción 1



- **Opción 2:**

File – Import Dataset – From (tipo de datos)

- **Opcion 3:**

Escribir funciones directamente e insertar directorio donde tenemos el archivo

Tipo de datos	Paquete	Función	Observacion
<i>csv</i>	readr	read_csv()	
<i>xls/xlsx</i>	readxl	read_excel()	Lee solo una hoja
	rio	import_list	Lee todo el libro de excel
<i>sas7bdat</i>	haven	read_sas()	
<i>dta</i>	haven	read_sata()	
	readstata13	read.dta13()	Convierte objetos de tipo 'labelled' automáticamente

Nota: Existen otros tipos de datos/paquetes, esta tabla muestra los más populares

Estructura:

Lado izquierdo asigno nombre = lado derecho leo mis datos. Ejemplo:

```
mis_datos = read_csv("mis_datos")
```

Nota La mayoría de las funciones para leer datos, convierten nuestros datos automáticamente en dataframes/tibbles, siempre debemos verificar la clase al final de leer los datos

Manejo de datos



- Dimensión:
`dim(mis_datos)`
- Número de columnas
`ncol(mis_datos)`
- Número de filas
`nrow(mis_datos)`
- Ver las primeras observaciones de nuestros datos
`head(mis_datos)`

Podemos aplicar funciones estadísticas a columnas de nuestros datos, ejemplo:

- Media:

```
mean(mis_datos$edad)
```

- Mediana

```
median(mis_datos$edad)
```

- Cuantiles

```
quantile(mis_datos$edad)
```

- Desviación estandar

```
sd(mis_datos$edad)
```

- Varias estadísticas al mismo tiempo

```
summary(mis_datos$edad)
```

Podemos contar observaciones en columnas categóricas, ejemplo:

```
table(mis_datos$sexo, exclude = NULL)
```

Nota: Recuerda que podemos verificar cómo usar una función con ayuda del panel help

Introducción a tidyverse

El *tidyverse* es una colección de paquetes que comparten la misma filosofía de diseño, gramática y estructuras de datos.



Los paquetes más famosos son:

- **dplyr**: manejo de datos
- **ggplot2**: creación de gráficos

Funciones básicas de dplyr

- *mutate()* añade nuevas variables en funcion de otras variables
- *select()* selecciona variables en base a sus nombres
- *filter()* selecciona casos en base a sus valores

Los pipes permiten tomar la salida de una función y enviarla directamente a la siguiente, lo cual es útil cuando necesita hacer muchas cosas con el mismo conjunto de datos

El símbolo que utilizamos es: `%>%`:

- *command-shift-M* en Mac
- *Ctrl-Shift-M* en Windows

Live coding

- P. Kuhnert & B. Venables, An Introduction to R: Software for Statistical Modeling & Computing
- Golemund, G., & Wickham, H. (2017). R for Data Science. O'Reilly Media.
<https://r4ds.had.co.nz/>
- *<https://www.tidyverse.org/>*
- *<https://dplyr.tidyverse.org/>*