

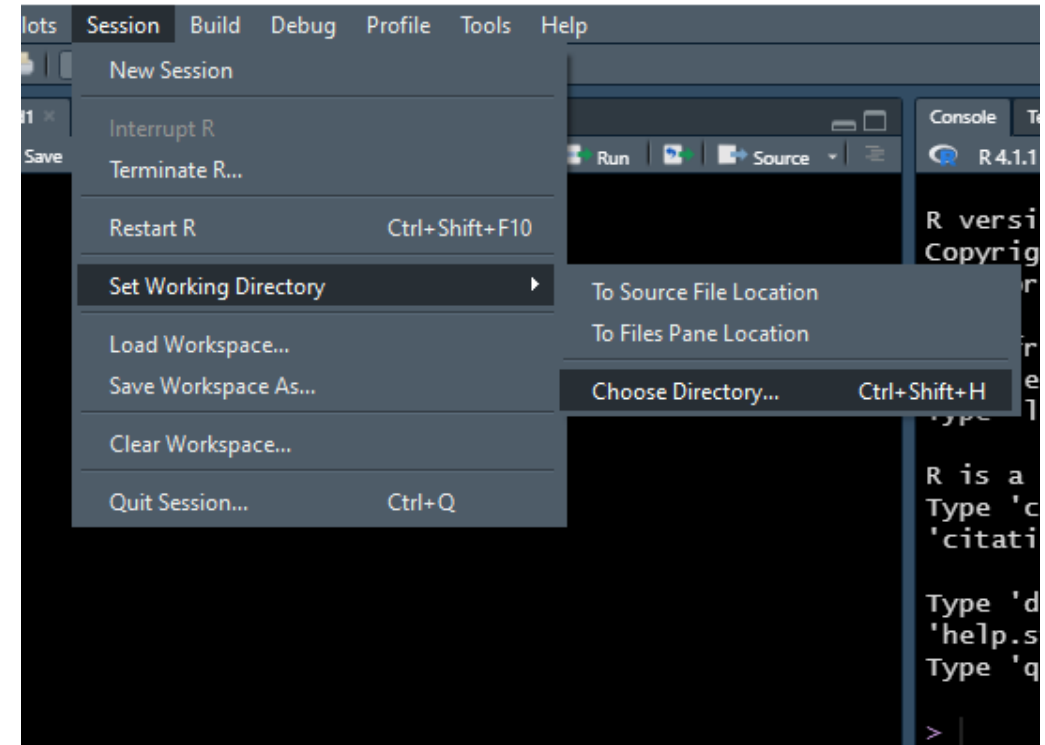
R Intermedio para Ciencias Sociales y Gestión Pública

SESIÓN N° 1: Cómo conseguir la data

Lo indispensable: el Directorio de trabajo

Para comenzar el trabajo con el R Studio lo primero que debemos hacer es decirle al programa cuál es la ubicación de nuestra carpeta de trabajo (*working directory*).

El *working directory* es el lugar en nuestra computadora (local) en el que están los archivos que vamos a utilizar durante nuestra sesión de trabajo. De la misma manera, es el lugar donde se encontrarán todos los documentos u objetos que vamos a producir en nuestro análisis (tablas, gráficos, bases de datos, entre otros).



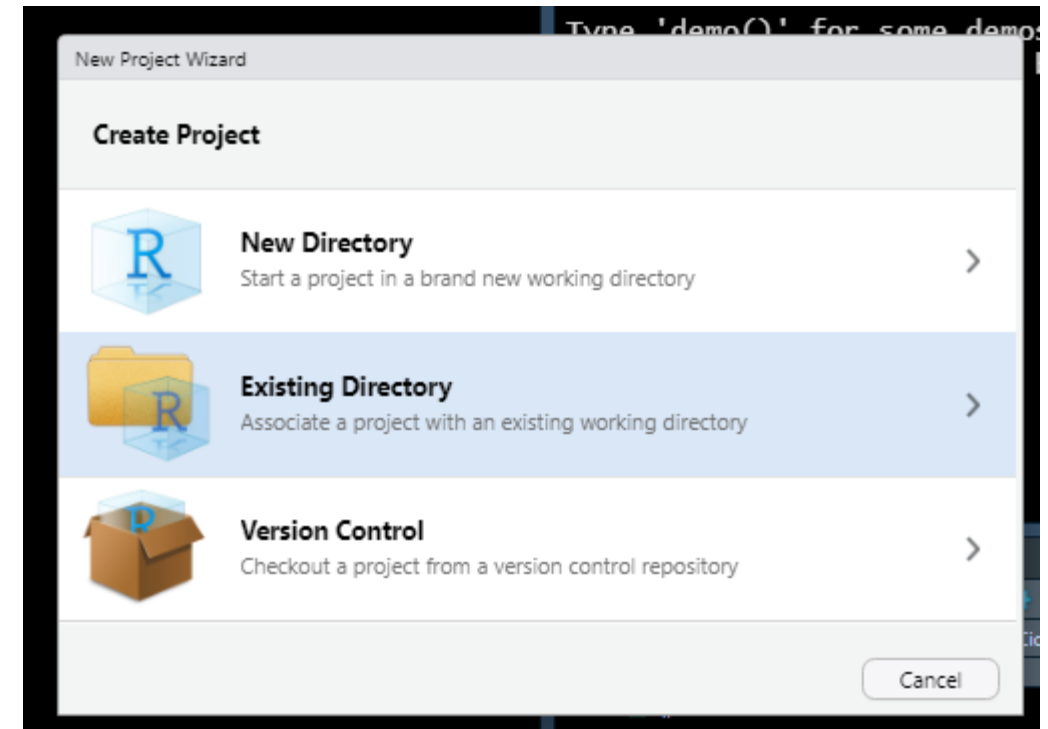
Lo indispensable: Proyectos de R

Lo más recomendable cuando se emprende una nueva tarea en el R es crear un **proyecto**.

Al crear un proyecto de R todos los ficheros (script de R, documento de Rmarkdown, etc.) quedan vinculados directamente al proyecto.

Ventajas:

- Es más ordenado pues consolida todos los outputs en un solo “lugar”.
- Guardar la *history* de todos los cálculos realizados.
- Hace más sencillo compartir la información.

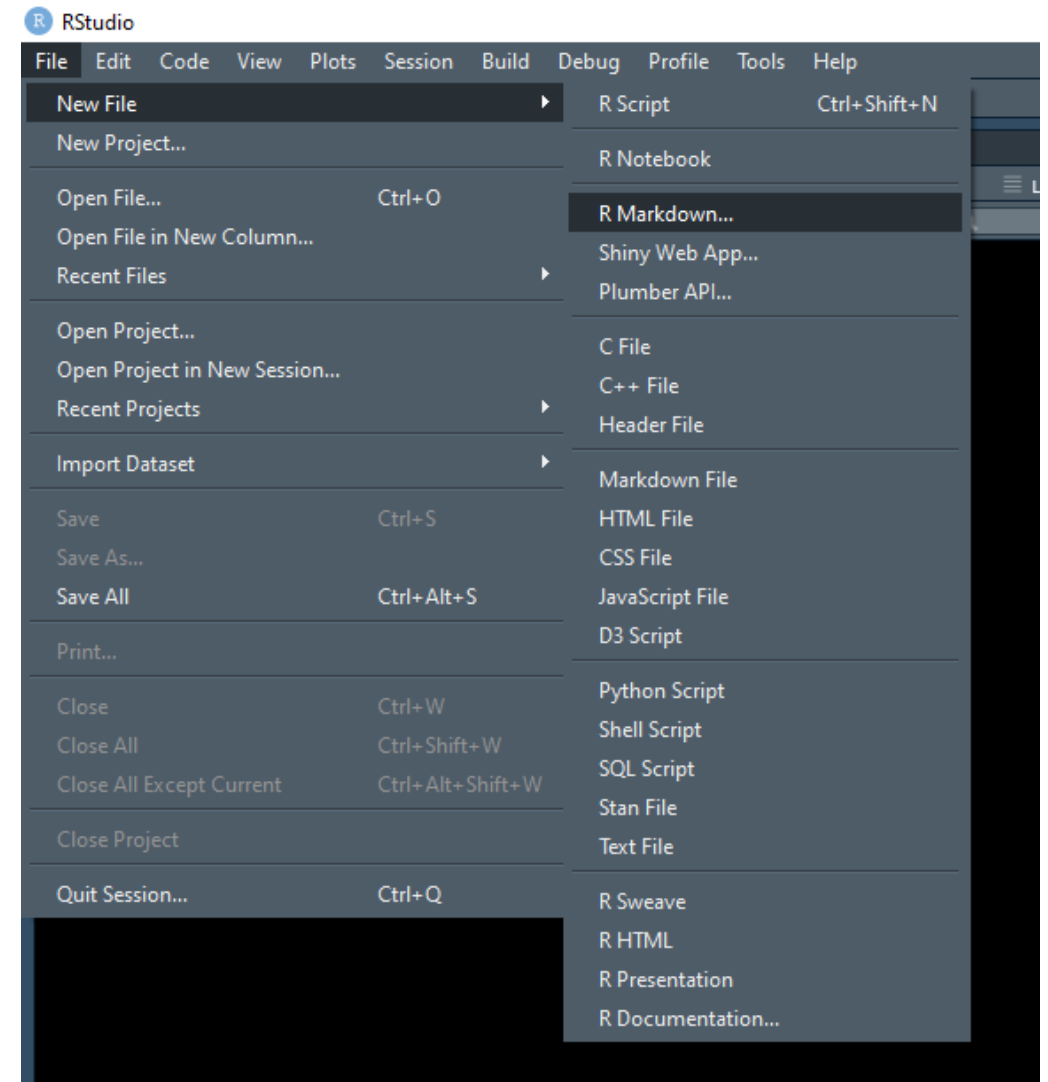


Lo indispensable: R Markdown

El R Markdown es una sintaxis de formato simple que nos permite elaborar informes de manera rápida y sencilla.

Recordar:

- Nos permite incluir texto sin necesidad que aparezca como comentario.
- Es necesario conocer lo básico de la redacción html (títulos, subtítulos, tablas, negrita, cursiva, numeración y viñetas).
- Incluir fragmentos de código (Chunks) que son ejecutados por el programa.
- PCs y TRABAJO FINAL se entregarán en reporte. NO EN SCRIPT



Lo indispensable: Estructuras y tipos de objetos

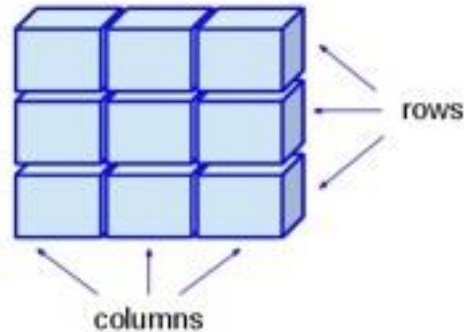
El R es un lenguaje de programación orientado a objetos.

TIPOS DE ESTRUCTURAS EN R

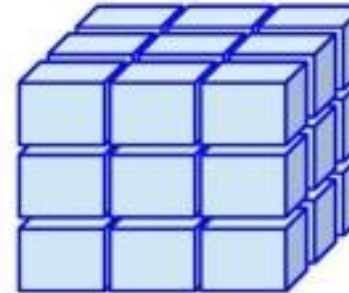
Vector



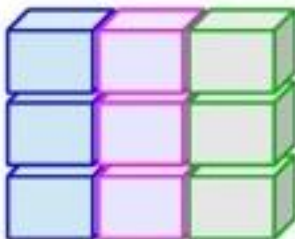
Matriz



Array

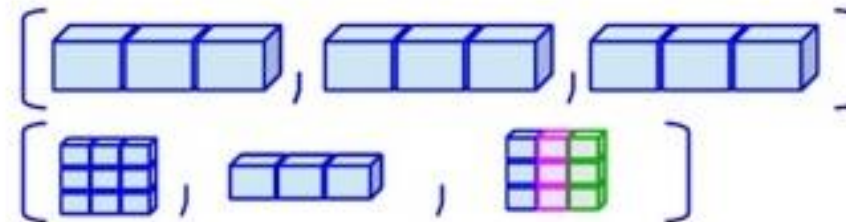


**Data frame
(tabla)**



**De distinto tipo*

Listas



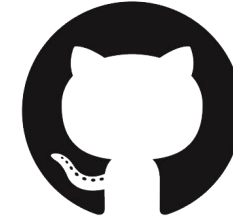
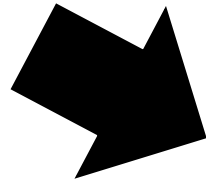
Lo indispensable: Estructuras y tipos de objetos

El R es un lenguaje de programación orientado a objetos.

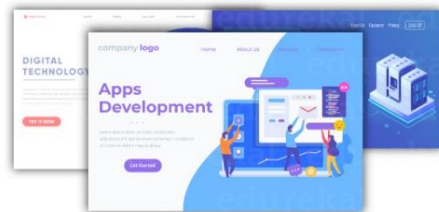
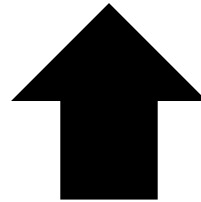
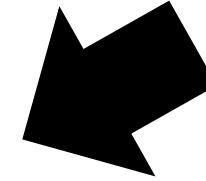
| Tipo de vector | Descripción |
|--------------------|--|
| Vector de factores | Vector que sirve para representar variables categóricas. Pudiendo ser factores nominales u ordinales. Ej: "costa", "sierra", "selva" |
| Vectores numéricos | Vector conformado por la concatenación de números. Ej: 1.5 , 2.7 , 3.1 |
| Matrices | Son hojas de datos homogéneos, es decir, de un sólo tipo. Tiene columnas y filas |
| Data frames | Son hojas de datos, estructuras similares a una matriz; sin embargo, a diferencia de estas pueden almacenar objetos de distintos tipo. Generalmente nuestras bases de datos son de este tipo (como una hoja de Excel que tiene variables numéricas, categóricas, etc). |
| Listas | Conjunto de objetos de distinta clase. Ej: "rojo", "1", "lápiz". Algunas veces las vamos a utilizar a la hora de hacer nuestros cálculos. Tener en cuenta. |

La clase de un objeto determinará de qué modo será tratado por el programa. El programa está dirigido a objetos y las funciones que están dentro de él también.

Nuevas fuentes de información



GitHub



Webpages



Web Scraping



Structured Data

edureka!



GitHub

Un “Drive” para programadores.

De la misma manera que el Google Drive o el Dropbox, el GitHub (creado el 2008) puede servir como un lugar de almacenamiento donde nosotros podemos guardar los documentos que vamos a utilizar en nuestra análisis: bases de datos, scripts, objetos producidos (tablas, gráficos), etc.

Sin embargo, la característica más fundamental del GitHub no es sólo su capacidad de almacenamiento, sino que es un espacio específicamente diseñado para desarrolladores de software. De esta manera, uno puede alojar su proyecto en el sitio web y puede gestionar diversas versiones utilizando el sistema de control de versión GIT.

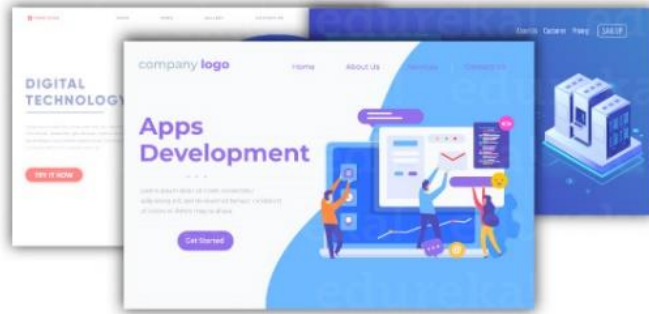


Asignación:

Cada alumno debe crear su cuenta de GitHub.

Puede usar el programa GitHub Desktop (tutorial en Paideia)

Web Scrapping



Webpages

Identificamos una página web en la cual se encuentren datos de nuestro interés.

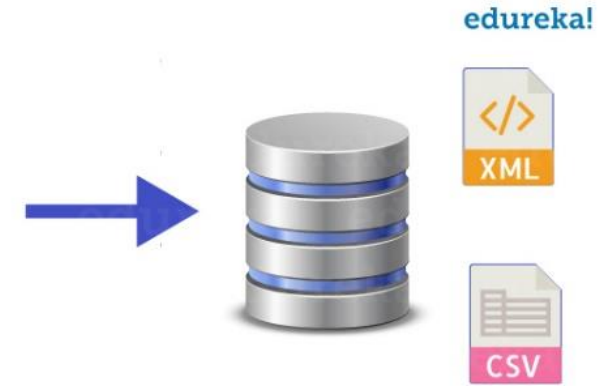
Estos pueden estar como tabla (en el mejor de los casos) o sueltos dentro de la estructura de la web.



Web Scrapping

Dependiendo de la estructura podemos aplicar algoritmos o funciones de extracción.

Una de los métodos más simples es la extracción de tablas html en páginas web **estáticas**.



Structured Data

Una vez extraída la información la mayoría de las veces se hace necesario la limpieza de la misma.

Luego de ello, recién podremos empezar a analizarla.

Otro tema clave! Funciones en R

Todo lo que queramos que ocurra en R debemos realizarlo con funciones, existentes o propias, de ahí la importancia de este tema.

¿Cuál es la estructura de una función?

En una función tenemos 3 tipos de elementos:

1.Argumentos (o valores de entrada).

2.Cuerpo: operaciones que han de realizarse. Se deben localizar entre corchetes "{}".

3.Resultado (o valores de salida): la última expresión que se ejecuta.

ESTRUCTURA:

```
Nombredelafunción <- function(argumentos){  
  OPERACIONES O CUERPO DE LA FUNCIÓN  
  return(resultado)  
}
```

EJEMPLO:

```
exp2 <- function(x){  
  z<-x*x  
  return(z)  
}
```

Recuerda: En la creación de la función se debe ser súper claro. No olvides colocar las llaves. Las funciones son clave durante el proceso de programación dado que pueden aligerar la carga de forma significativa y también nos ayuda a evitar ser repetitivos.

Recuerda guardar la data que has trabajado...

Para exportar un archivo que tengamos en nuestra sesión sólo debemos escribir la siguiente línea de comando:

```
export(nombre_del_objeto, "archivofinal.csv")
```

Recuerda que si gustas también puedes **convertir** un archivo de un formato a otro. Por ejemplo, si tengo la base de datos de trabajadores.sav en mi carpeta de trabajo y deseo convertirla a formato .rda, sólo tengo que escribir el siguiente comando:

```
convert("trabajadores.sav", "trabajadores.rda")
```

RECUERDA LOS TIPOS DE FORMATO: Excel (.xls, .csv), SPSS (.sav), STATA (.dta) o los del propio R (.rda)