

Ciencia de Datos



Ciencia de Datos

Módulos y Agenda

Instructores



Módulos: Ciencia de Datos - 4:00 a 6:30 pm

Módulos		Horas	Fechas	Instructor
I.	Introducción a la ciencia de datos	20	Del 11,12, 13, 14 y 18, 19, 20, 21 de octubre 2021	Act. Arturo Téllez Cortés
II.	Herramientas computacionales para el manejo de datos: R			
III.	Herramientas computacionales para el manejo de datos: Python	20	Del 25, 26, 27, 28 de octubre y 1, 3, 4, 5 de noviembre 2021	Act. Arturo Téllez Cortés
IV.	Herramientas analíticas: Análisis exploratorio de datos			
V.	Modelos analíticos de minería de datos I: Modelo lineales	20	Del 8, 9,10, 11 y 16, 17, 18, 19 de noviembre 2021	Act. José Eduardo Lozas Galicia
VI.	Modelos analíticos de minería de datos II: Clasificación y Redes Neuronales			
VII.	Aplicaciones: Predicción, Minería de datos y textos y otras	20	Del 22, 23, 24, 25 de noviembre y 29, 30 noviembre, 1 y 2 de diciembre de 2021	Mtro. Javier Juárez Martínez
VIII.	Desarrollo de un proyecto: analiza tus propios datos			

Incluye temas relacionados con la gestión de datos, la estadística, la minería de datos y el aprendizaje de máquina.

La ciencia de datos permite extraer diferentes tipos de patrones en diferentes contextos, por ejemplo:

Segmentación de clientes: extraer patrones que permiten identificar grupos de clientes que muestran conducta o gustos similares.

o Reglas de asociación: patrones que identifican productos que son comprados conjuntamente,

o Identificación de anomalías: patrones que muestran actividad fraudulenta.

Ciencia de Datos

I.- Introducción a la ciencia de datos

II.- Herramientas computacionales para el manejo de datos: R

Presentación de Participantes



Presentación de Participantes



Antes de Iniciar, vamos a conocernos un poco

 Nombre y Formación

 Antigüedad en CNBV

 Puesto

 Funciones en CNBV

 Área

 Expectativas del Curso

Ciencia de Datos

I.- Introducción a la ciencia de datos

II.- Herramientas computacionales para el manejo de datos: R

Objetivo Particular



El participante aprenderá los conocimientos básicos de la ciencia de datos para despertar el interés en este campo de aplicación moderna en el sector financiero y la industria

Ciencia de Datos

I.- Introducción a la ciencia de datos

II.- Herramientas computacionales para el manejo de datos: R

Temario



■ ■ ■ ■ ■ I.- Introducción a la ciencia de datos

1. Definición de ciencia de datos y temas afines (aprendizaje de máquina, minería de datos, big data, aprendizaje estadístico, etcétera).
2. Ecosistema de la ciencia de datos.
3. Proceso para el análisis de datos.
4. Ética y privacidad

1. Instalación de R, R studio y otras herramientas de apoyo. Filosofía de R.
2. Estructuras de datos y su manipulación en R.
3. Elementos de programación.
4. Modelación en R.
5. Reportes en R

Ciencia de Datos

I.- Introducción a la ciencia de datos

II.- Herramientas computacionales para el manejo de datos: R

Evaluación



Evaluación Diagnóstica



Evaluación Final:

✓ Asistencia y puntualidad	20%
✓ Participación	30%
✓ Examen final	50%
✓ Total	100%

Asistencia:

- ✓ Se requiere el 80% para tener derecho a Examen y Constancia.

Ciencia de Datos

I.- Introducción a la ciencia de datos

II.- Herramientas computacionales para el manejo de datos: R



Ciencia de Datos

I.- Introducción a la ciencia de datos



Ciencia de Datos

I.- Introducción a la ciencia de datos

1.- Definición de ciencia de datos y temas afines (aprendizaje de máquina, minería de datos, big data, aprendizaje estadístico)

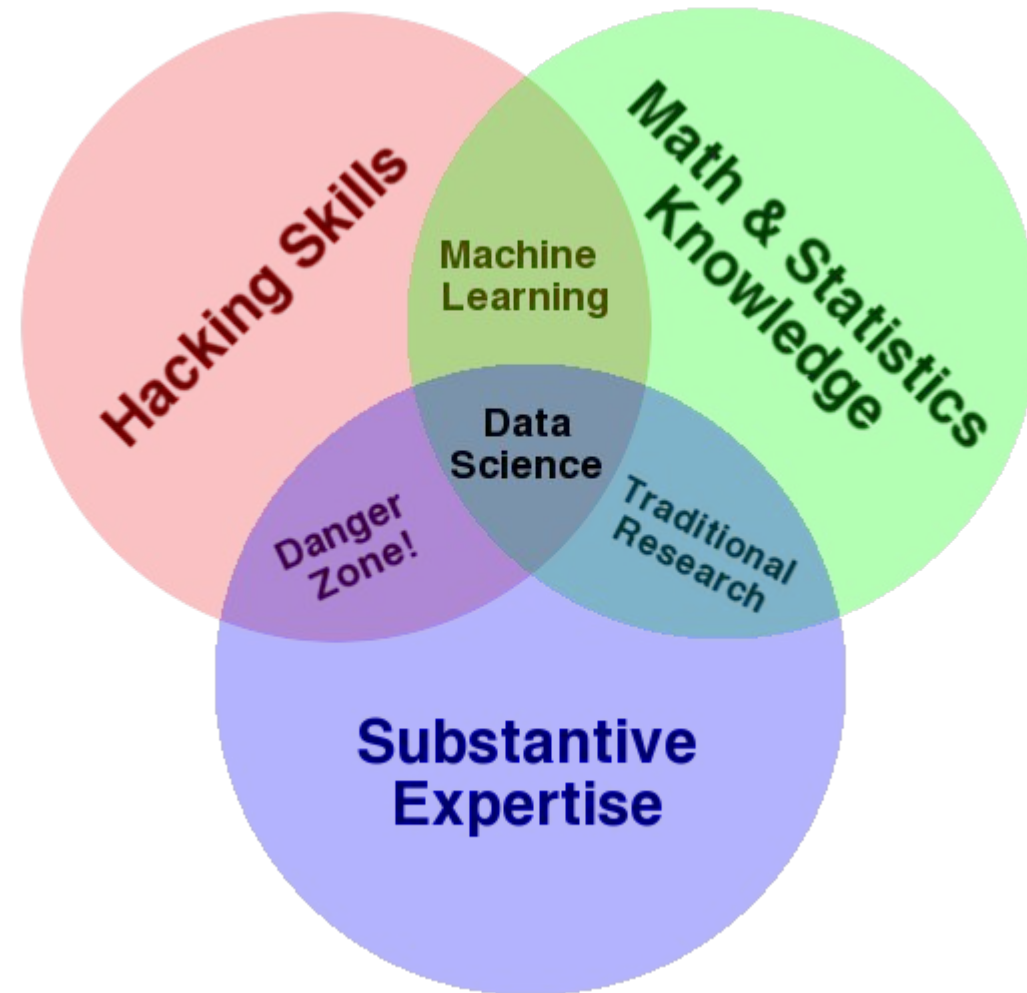


Definición de ciencia de datos y temas afines

La ciencia de datos

La ciencia de datos (*data science*) es una disciplina fascinante que te permite convertir datos sin procesar en entendimiento, comprensión y conocimiento.

<https://es.r4ds.hadley.nz/introducci%C3%B3n.html>



Drew Conway

Definición de ciencia de datos y temas afines

● **Data Science**
Término de búsqueda

● **Ciencia de datos**
Campo de estudio

+ Añadir comparación

México ▼

Últimos 5 años ▼

Todas las categorías ▼

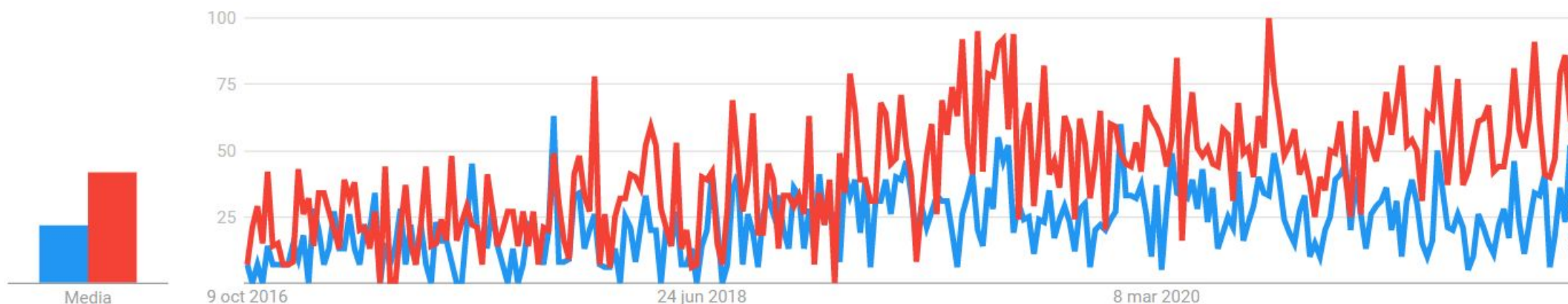
Búsqueda web ▼

! **Nota:** Esta comparación contiene temas y términos de búsqueda, los cuales se miden de forma diferente.

[MÁS INFORMACIÓN](#)

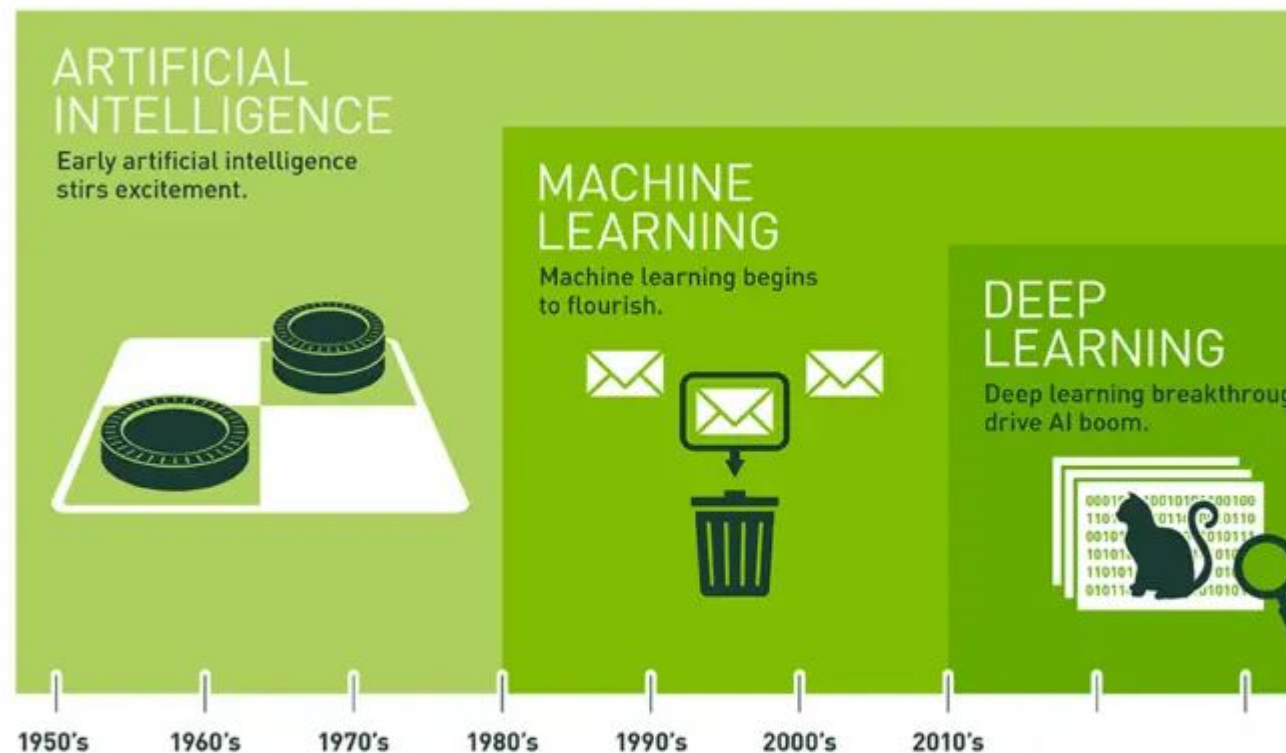
Interés a lo largo del tiempo ?

https://trends.google.es/trends/explore?date=today%205-y&geo=MX&q=Data%20Science,%2Fm%2F0jt3_q3



Definición de ciencia de datos y temas afines

Inteligencia Artificial

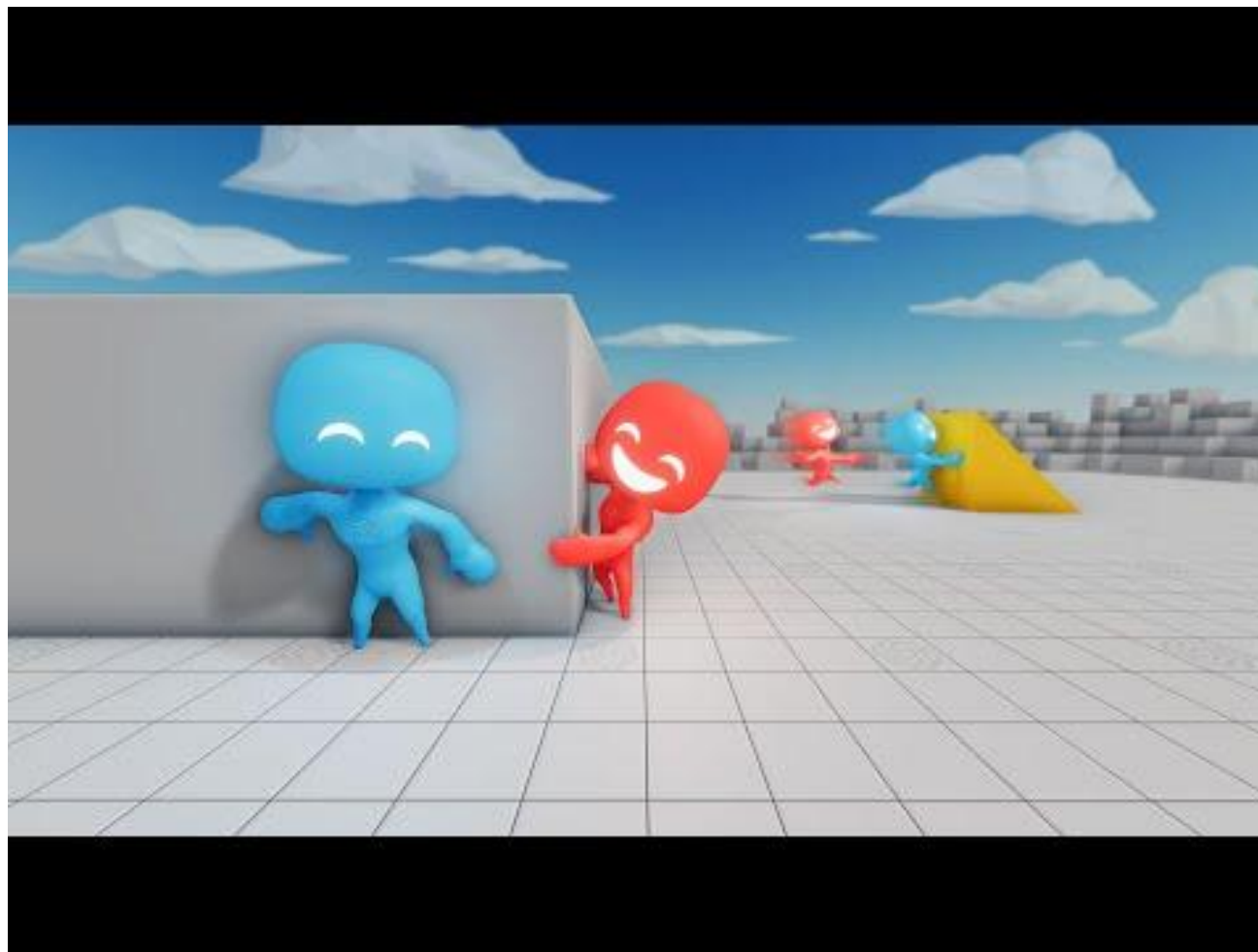


Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

<https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>

Definición de ciencia de datos y temas afines

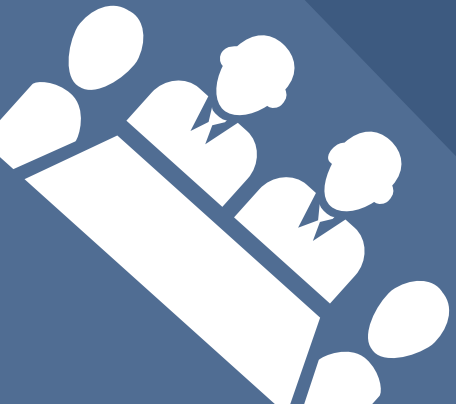
Aprendizaje por refuerzo



Ciencia de Datos

1.- Introducción a la ciencia de datos

2.- Ecosistema de la ciencia de datos



Ecosistema de la ciencia de datos

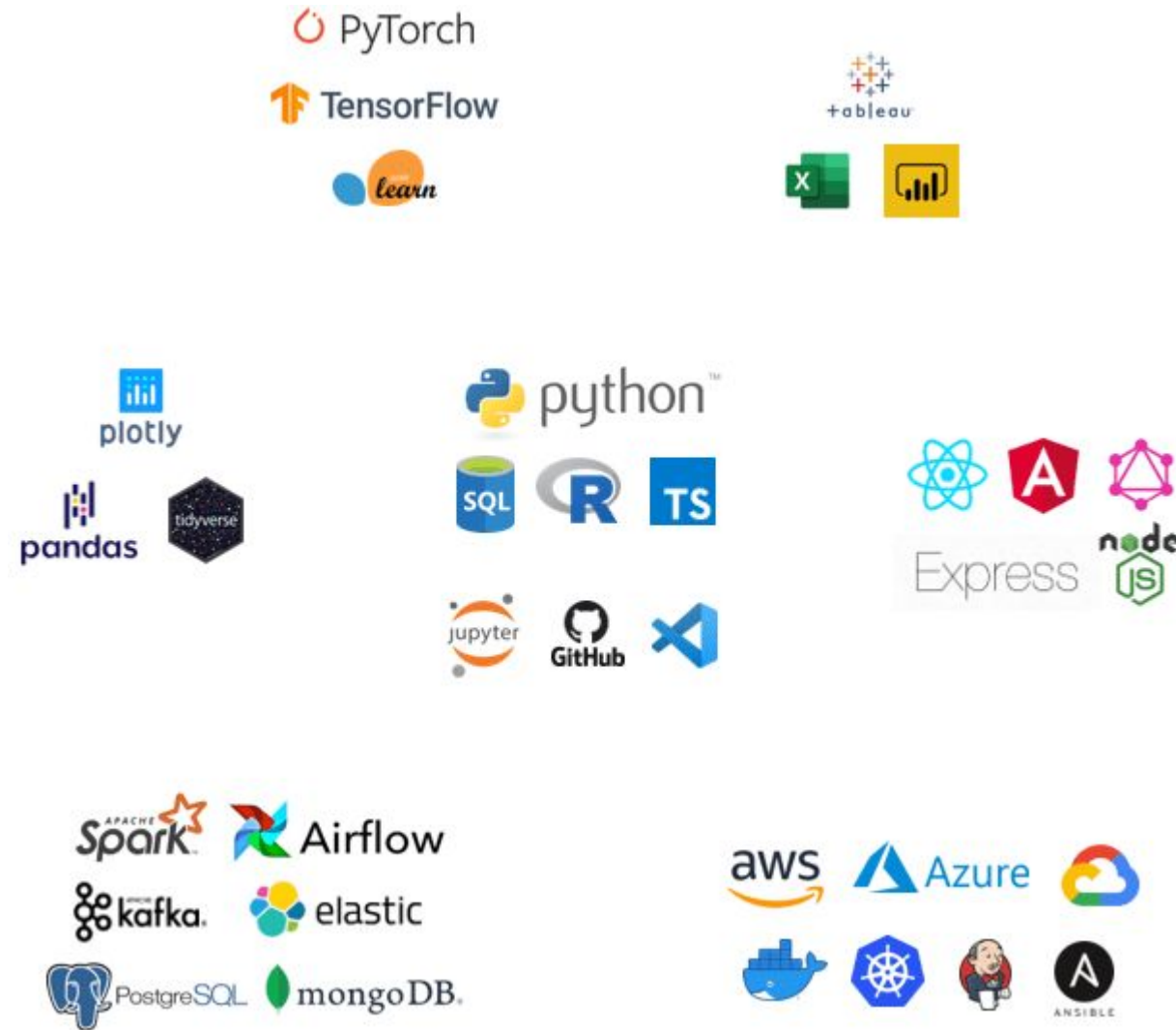
La ciencia de datos

DATA SCIENCE TOOL	◄ % OF RESPONDENTS USING THE TOOL ►
Python	86.7
SQL	42.1
R	23.9
C++	21.4
Java	18.8
C	18.5
JavaScript	16.7
MATLAB	12.4
Other	10.9
Bash	9.9

<https://research.aimultiple.com/data-science-tools/>

Ecosistema de la ciencia de datos

Herramientas



Ciencia de Datos

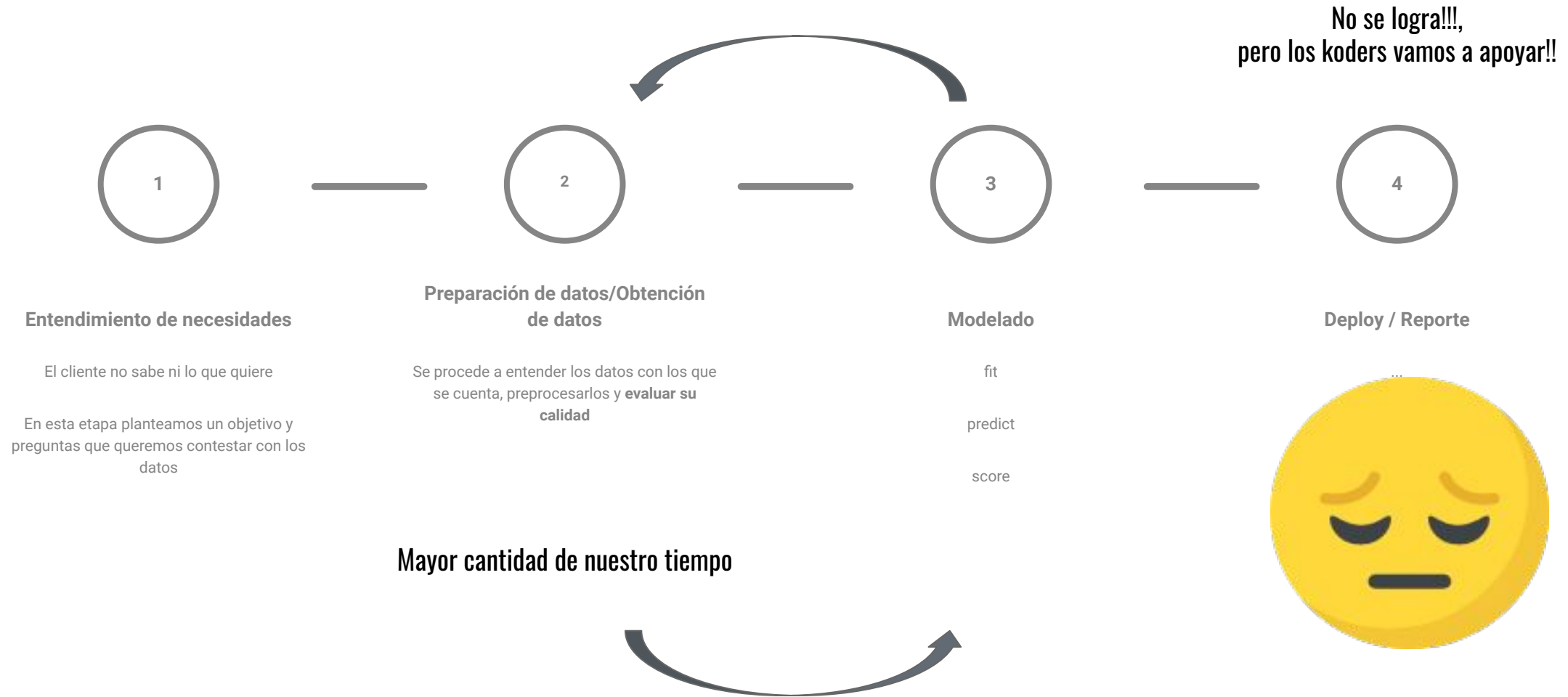
I.- Introducción a la ciencia de datos

3.- Proceso para el análisis de datos



Proceso de análisis de datos

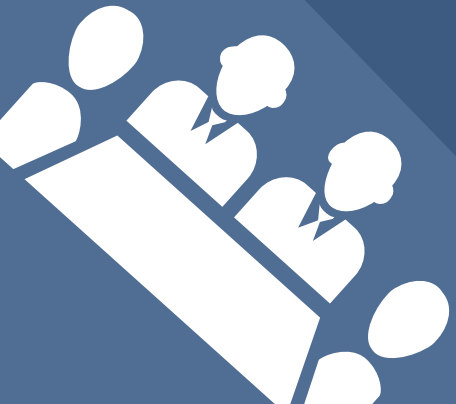
Herramientas



Ciencia de Datos

I.- Introducción a la ciencia de datos

4.- Ética y privacidad



La tecnología no es el problema, sino cómo la instruimos y para qué la utilizamos.

La ética debe estar presente en todo el ciclo de vida del dato: captación, gestión, privacidad y uso

¿Para qué se van usar esos datos?

■ ■ ■ ■ ■ Ética y privacidad

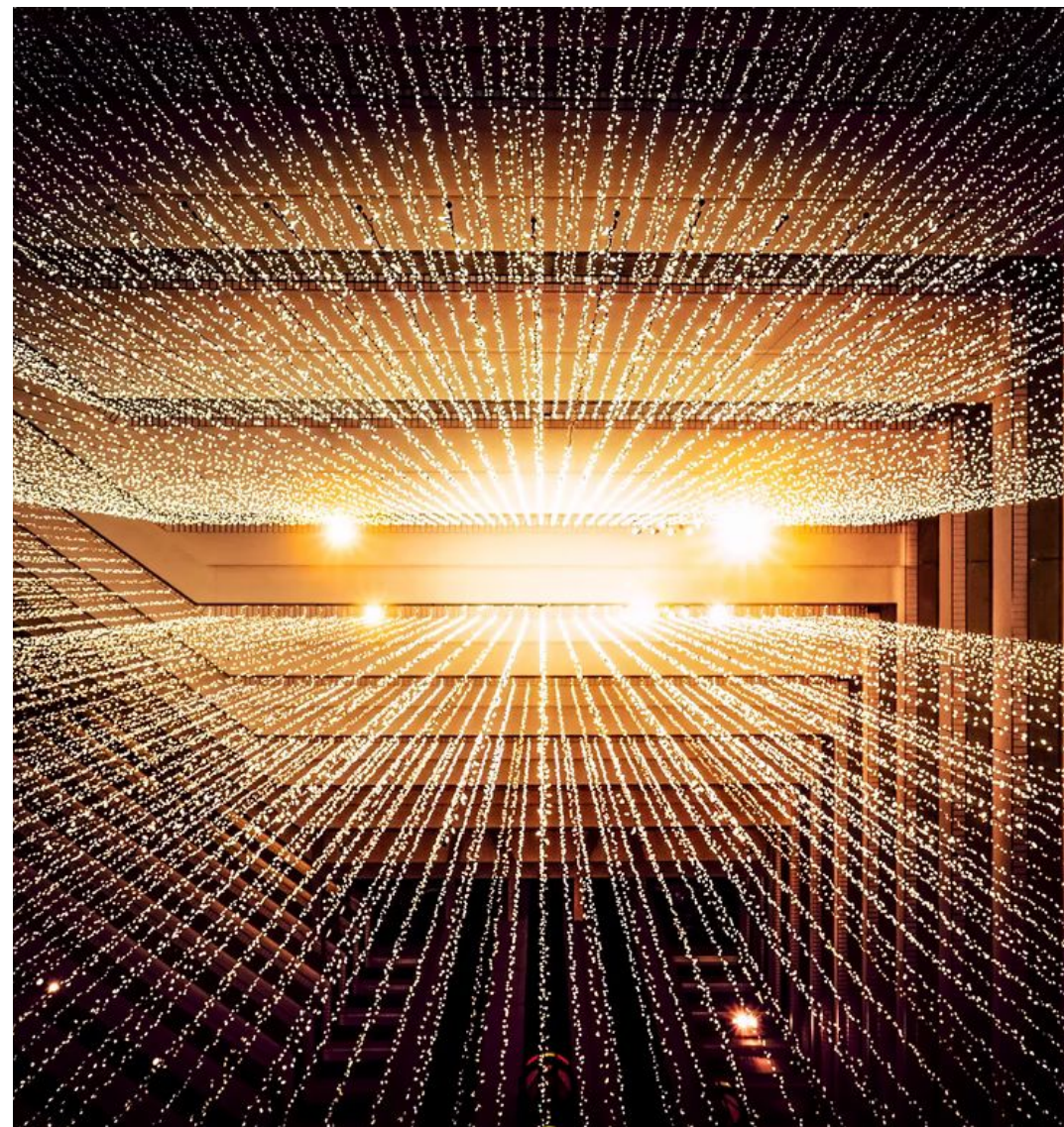
“Las organizaciones que actúen con ética y transparencia en el uso de los datos tendrán una valoración positiva de la sociedad. Y al contrario, quienes no se rijan por esos parámetros sufrirán una enorme pérdida en su reputación”

“en la era digital, la **confianza en una organización** no se mide solo por su marca, líderes, valores o empleados. También, y cada vez más, por **su forma de gobernar y gestionar los datos, los algoritmos**, las máquinas...”.

Eva García San Luis, socia responsable de Data & Analytics de KPMG.

■ ■ ■ ■ ■ Ética y privacidad

Según una reciente encuesta de KPMG entre 2.190 directivos de diferentes partes del mundo, el **92% se muestra preocupado** por el **impacto** que puede tener en la **reputación corporativa** la falta de confianza en la gestión que hace de los datos.



Cambridge Analytica: la multa récord que deberá pagar Facebook por la forma en que manejó los datos de 87 millones de usuarios

Redacción
BBC News Mundo

24 julio 2019

<https://www.bbc.com/mundo/noticias-49093124>

■ ■ ■ ■ ■ Ética y privacidad

Es cuestión de cómo se ve

en India, el 78% de los consumidores ve bien que los taxis utilicen datos de geolocalización para proponer rutas a los clientes, algo que en Dinamarca no gusta al mismo porcentaje.



■ ■ ■ ■ ■ Ética y privacidad

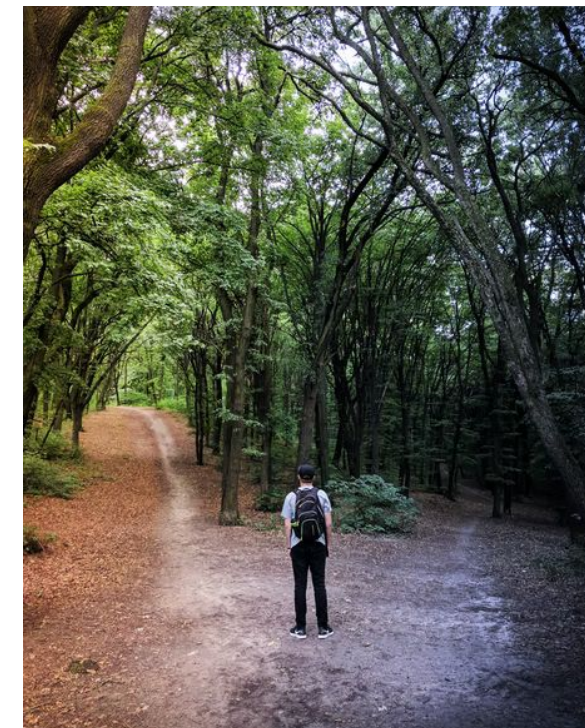
En resumen

En torno a la Inteligencia Artificial, el debate es **cómo lograr que las máquinas inteligentes estén guiadas por la ética y orientadas hacia el bien común**

Los **pilares** sobre los que debe **articularse esa confianza** son **calidad y precisión de los datos, seguridad y protección** de los mismos e integridad en su uso.

La clave es **comunicación, honestidad y transparencia**; explicarles cómo y para qué se utilizan sus datos y las ventajas que les reportan a ellos”,

La ética del Data, <https://www.tendencias.kpmg.es/2018/04/etica-big-data/>



Vamos a jugar
<https://www.moralmachine.net/hl/es>

Estrategia de Inteligencia Artificial MX 2018

México se convierte en uno de los primeros 10 países en contar con una estrategia para avanzar en materia de Inteligencia Artificial.

Por Enrique Zapata | *Director General de Datos Abiertos*

- Se llevó a cabo la presentación del estudio “En miras hacia una Estrategia de Inteligencia Artificial (IA) en México: Aprovechando la Revolución de IA”, disponible al público la primera semana de abril.
- Durante el evento se presentaron las acciones en materia de **Inteligencia Artificial MX para avanzar en las recomendaciones de corto plazo.**

■ ■ ■ ■ ■ Ética y privacidad

La Ley de Protección de Datos en México

México cuenta con una ley que regula el tratamiento de los datos personales por parte de empresas del sector privado desde el 5 de julio de 2010, esa ley se llama Ley Federal de Protección de Datos Personales en Posesión de los Particulares o Ley de Protección de Datos.

Su aplicación determina que se evite que los datos personales sean utilizados indebidamente, que se respeten los derechos de los dueños de los datos y que se garantice una expectativa razonable de privacidad. Quienes traten datos personales deben tomar en cuenta las guías y documentos emitidos por el Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales ("INAI").

<https://www.salesforce.com/mx/blog/2020/11/ley-de-proteccion-de-datos-personales.html>



Ciencia de Datos

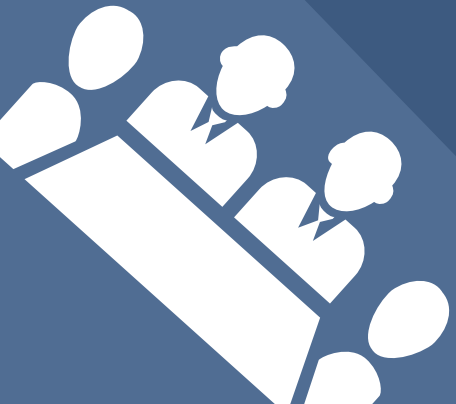
I.- Introducción a la ciencia de datos

II.- Herramientas computacionales para el manejo de datos: R



Ciencia de Datos

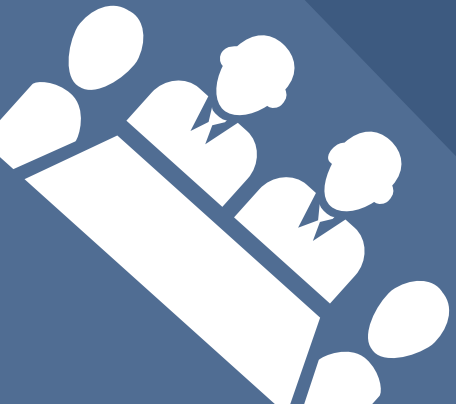
II.- Herramientas computacionales para el manejo de datos: R



Ciencia de Datos

II.- Herramientas computacionales para el manejo de datos: R

1.- Instalación de R, Rstudio y otras herramientas de apoyo. Filosofía de R



■ ■ ■ ■ ■ Lenguaje de programación R

Instalación

RStudio Desktop 2021.09.0+351 - [Release Notes](#)

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:



DOWNLOAD RSTUDIO FOR WINDOWS

2021.09.0+351 | 156.88MB

Requires Windows 10 (64-bit)

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



■ ■ ■ ■ ■ Lenguaje de programación R

Filosofía de R

“Buscamos que los usuarios puedan iniciar en un entorno interactivo, en el que no se vean, conscientemente, a ellos mismos como programadores. Conforme sus necesidades sean más claras y su complejidad se incremente, deberían gradualmente poder profundizar en la programación, es cuando los aspectos del lenguaje y el sistema se vuelven más importantes.”

a John Chambers (Chambers and Hastie [1991])

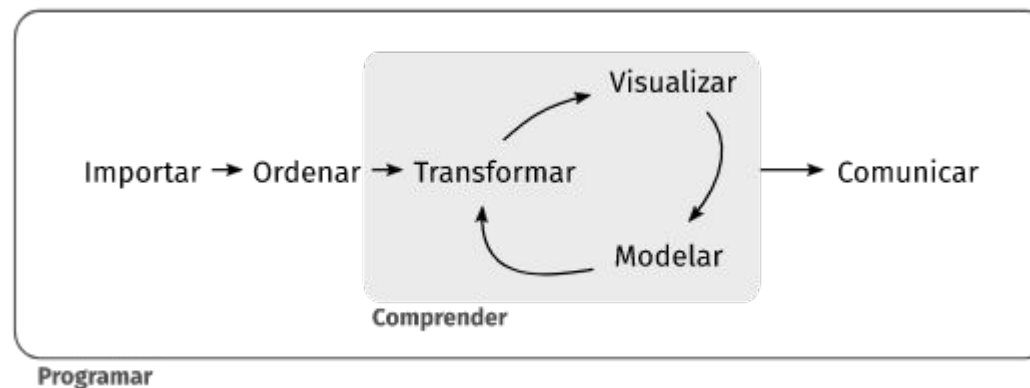


Lenguaje de programación R

¿Por qué R?

R es un lenguaje de programación de propósito estadístico, nos permite:

- Cargar datos
- Prepararlos (Ordenarlos)
- Transformarlos
- Visualizarlos
- Modelarlos
- Comunicarlos



<https://es.r4ds.hadley.nz/index.html>

Ciencia de Datos

II.- Herramientas computacionales para el manejo de datos: R

2.- Estructuras de datos y su manipulación en R



Estructuras de datos y su manipulación en R.

Vectores

R tiene cinco clases básicas de vectores:

- character (letras)
- numeric (números reales)
- integer (números enteros)
- complex (números complejos)
- logical (verdadero/falso o `TRUE (T)` / `FALSE (F)`)

Lo más importante es que los vectores sólo pueden contener elementos de la misma clase. Por ejemplo, no es posible tener un vector de la siguiente forma: `c(1, "perro", FALSE, "gato", 1.5)` pero sí es posible tenerlo de esta forma: `c("1", "perro", "FALSE", "gato", "1.5")`. En este caso, aunque bastante inconveniente, todos los elementos son de tipo `character`. R automáticamente fuerza el resultado para no obtener errores (en un proceso llamado coerción). En nuestro ejemplo:

Coerción

```
prueba <- c(1, "perro", FALSE, "gato", 1.5)
```

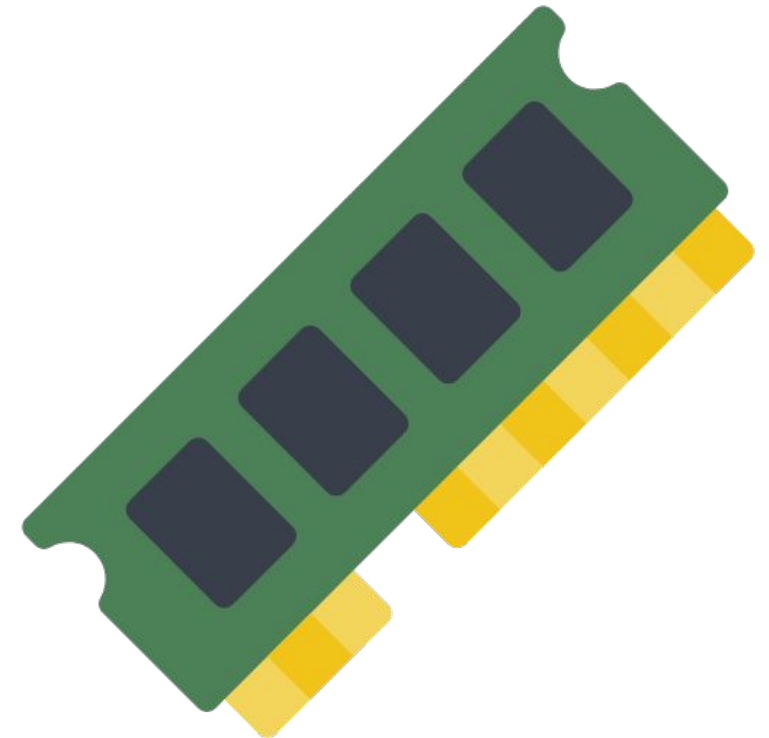
Otras clases de objetos incluyen listas (`list`), matrices (`matrix`), tablas de datos (`data.frame`) y modelos (por ejemplo `lm` para modelos lineales)

¿Qué es una variable?

Es un sitio donde guardamos una determinada información, y se les clasificará dependiendo el tipo de información que almacenen.

Características de las variables:

- Case Sensitive
- No se pueden usar como nombres de variables palabras reservadas como: if, else, for
- No se pueden incluir en nombre de variables con tilde o con ñ
- No PUEDEN INICIAR CON un número o contener “_”
- **La información no persiste al cerrar la sesión con el interprete**



■ ■ ■ ■ ■ Estructuras de datos y su manipulación en R.

Ciclos

while

```
while(TRUE){
```



for

```
for(i in 1:10) {
```



repeat

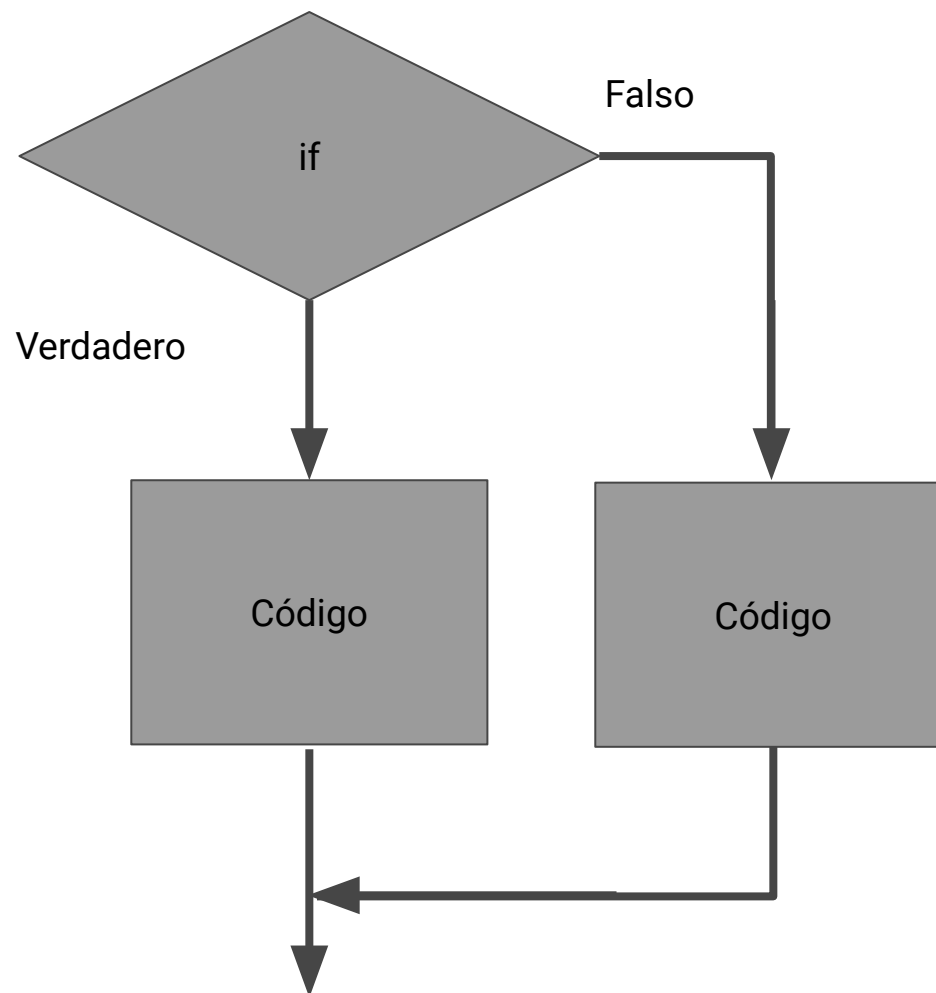
```
repeat({
```



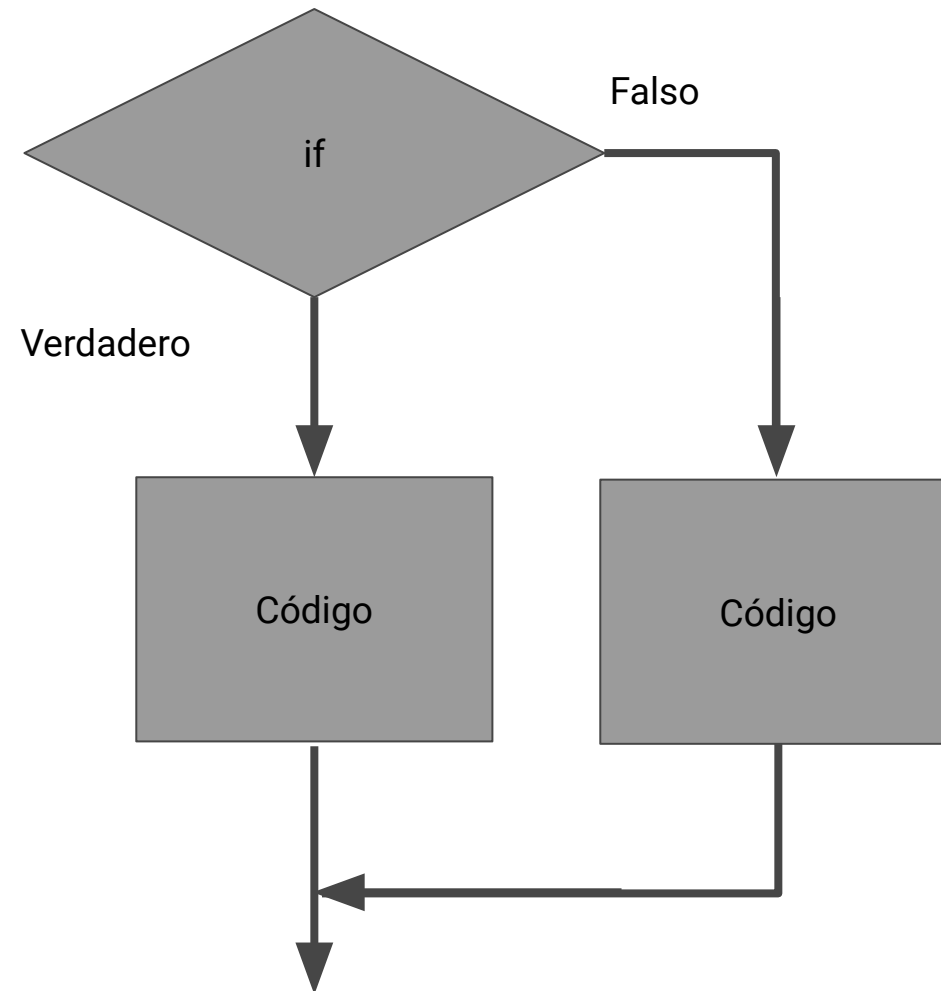
Estructuras de datos y su manipulación en R.

Condicionales

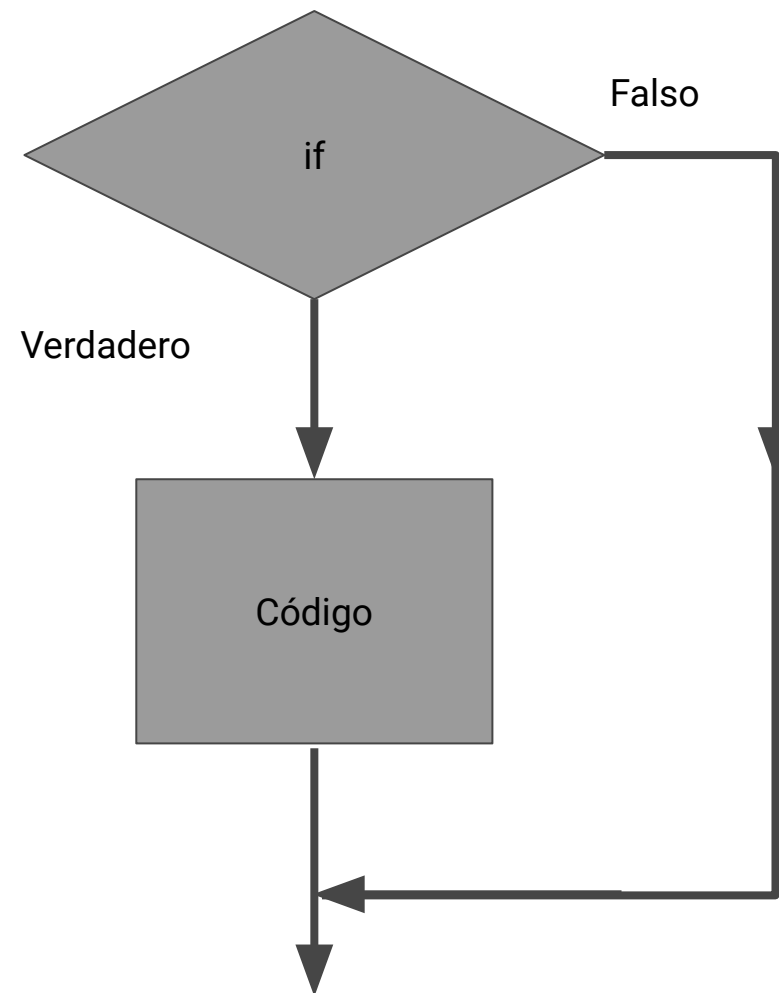
```
if(condicional1){  
  #Código  
}else if(condicional2){  
  #Código  
  ...  
}else{  
  #Código  
}
```



```
if(condicional1){  
    #Código  
}else{  
    #Código  
}
```



```
if(condicional1){  
  #Código  
}
```



■ ■ ■ ■ ■ Estructuras de datos y su manipulación en R.

Funciones

```
funcion_nombre <- function(parametros){  
  código  
}
```

```
1  
2 calculadora <- function(tipo, a, b){  
3   if(tipo == "suma"){  
4     return(x + y)  
5   }else if(tipo == "resta"){  
6     return(x - y)  
7   }else if(tipo == "multiplicacion"){  
8     return(x * y)  
9   }else if(tipo == "division"){  
10    return(x / y)  
11  }else{  
12    return(NA)  
13  }  
14  
15 }
```