

Importación y Tratamiento de datos

Flor Naomi Aguilar Portilla

25/10/2021

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Vamos a utilizar como ejemplo: **iris**, que es una matriz de datos precargada en R.

Abrir matriz de datos.

```
{r, warning=FALSE, message=FALSE} library(datasets) data(iris) {r, warning=FALSE, message=FALSE}
```

Exploración de los datos iris.

1.- Dimensión de la matriz.

```
dim(iris)
```

2.- Nombre de las columnas.

```
colnames(iris)
```

```
names(iris)
```

3.- Clase a la que pertenece la matriz de datos.

```
class(iris)
```

4.- Estructura interna.

```
str(iris)
```

5.- Observación de una variable específica.

```
iris$Species
```

6.- Visualización de tabla.

```
View(iris)
```

7.- Estadística descriptiva básica.

```
summary(iris)
```

Datos faltantes.

1.- Búsqueda de datos faltantes.

```
anyNA(iris)
```

Nota: Se le pregunta a R si hay datos faltantes (**NA**), R me responde con **FALSE** en caso de NO HABER NA's y **TRUE** en caso de HABER NA's.

2.- Suma de datos faltantes.

```
sum(is.na(iris))
```

3.- Librería **mice**.

a) Instalar el paquete **mice**

b) Función **md.pattern(iris)**, pero no me compila en el pdf, así que voy a insertar el gráfico de NA's como imagen. Nota: Se activó la función desde un script sencillo.

Datos atípicos.

a) Detección. Se detectan con el gráfico boxplot.

```
bx1<-boxplot(iris)
bx1
```

b) filtrado de variables para realizar el boxplot

```
bx2<-boxplot(iris[,c(1:4)], col = "cornsilk1")
bx2
```