

3. Exploración general de una BD

Notas varias, opiniones y observaciones - Curso de Luis Alejandro Carrera

Patricio Moracho

23/2/2020

```
#####  
## Exploraci?n de una base de datos ##  
#####
```

```
## La exploraci?n de una base de datos es el primer acercamiento que tenemos con nuestra  
## informaci?n. Esta actividad implica fundamentalmente observar las caracter?sticas generales  
## de nuestros datos como por ejemplo: n?mero de observaciones, n?mero de variables, nombres de  
## las variables, tipos de variables y observaciones perdidas (NA).
```

```
## La estructura general de la base de datos se obtiene con la funci?n str()  
str(mtcars)
```

`str()` viene de “structure” y es una función polimórfica ya que aplica a una enorme cantidad de objetos distintos. Vendría bien mencionar el concepto de datos estructurado y no estrucutrados, que ambos se usan todo el tiempo para generar información valiosa. Imagino que tu curso igual esta orientado básicamente a los estructurados.

```
## La funci?n regresa n?mero de observaciones, n?mero de variables y tipo de las variables  
## Otros comandos ?tiles son:  
dim(mtcars)           # Muestra el n?mero de renglones y columnas de una base de datos  
glimpse(mtcars)       # Funci?n de tidyverse. Vista de las variables como renglones  
head(mtcars, n); tail(mtcars,n)  # Despliega las primeras y ?ltimas observaciones  
                        # El segundo argumento de estas funciones es un entero  
nrow(mtcars); ncol(mtcars) # N?mero de renglones y n?mero de columnas  
names(mtcars)         # Nombres de variables (columnas)  
row.names(mtcars)     # Nombres de renglones (en caso de que existan)  
edit(mtcars)          # Permite la exploraci?n de una BD  
View(mtcars)          # Permite visualizar en formato data.frame cualquier base de datos
```

`dim()` es otra función polimórfica, lo que retorna son las dimensiones de un objeto, en el caso de un `data.frame` o una `matriz`, efectivamente es el número de filas y columnas. Un comentario sobre `edit()`, esta función sirve para editar (modificar) y generar nuevas copias de objetos, puede ser algo confusa por que varía en función de cada objeto:

Dentro de esta línea no olvidaría mencionar a `summary(mtcars)`

```
## Con la librer?a "inspectdf" se pueden realizar distintos tipos de exploraci?n  
## Algunas de las funciones m?s importantes son:  
library(inspectdf)  
inspect_types(mtcars)      # Nos presenta un resumen general de las variables (tipos)
```

```

inspect_cat(mtcars)      # Presenta un resumen general de cada variable categ?rica
                        # cnt indica el n?mero de valores ?nicos, common el valor m?s frecuente,
                        # common_pcmt el porcentaje de ocurrencia del valor m?s frecuente
inspect_imb(mtcars)      # Ordena las variables categ?ricas seg?n sus valores m?s comunes
inspect_num(mtcars)      # Presenta un resumen general de cada variable num?rica
inspect_na(mtcars)       # Muestra el conteo de NA por variable
inspect_cor(mtcars)      # Examina correlaciones por pares para las variables num?ricas

```

Buena idea presentar este paquete, es muy pr?ctico para esats tareas.

```

## Ejercicio: Examinar la base de datos "flights" del paquete "nycflights13" empleando las funciones
## vistas anteriormente.
## 1. ?Cu?ntas observaciones y cu?ntas variables tiene la base de datos?
## 2. ?Los nombres de las variables son claros y ?tiles?
## 3. ?Cu?ntas variables son num?ricas y cu?ntas son categ?ricas?
## 4. ?Existen registros perdidos?
## 5. ?Cu?les variables podr?an modificarse para que nos proporcionen informaci?n ?til?

install.packages("nycflights13")
library(nycflights13)
View(flights)

```