

Introducción a dplyr

Armando Ocampo

dplyr

dplyr es una librería que nos permite manipular datos, ordenarlos y extraer información necesaria. Esta paquetería no forma parte de las paqueterías base de R. Por lo que tenemos que descargarla.

```
install.packages("dplyr")
```

Una vez instalada, la llamaremos para comenzar a trabajar.

```
library(dplyr)
```

dplyr pertenece a un conjunto de paqueterías llamado tidyverse, el cual nos permite manipular información. Es bueno que sepas esto, no obstante, el día de hoy solo trabajaremos con la paquetería mencionada.

Las funciones principales de dplyr son `filter()`, `arrange()`, `mutate()`, `group_by()` y `summarise()`. También se conocen como verbos. Además existe un elemento denominado pipe "`%>%`" que nos permite unir dos elementos y trabajar con ellos.

filter()

El primer verbo que utilizaremos se llama `filter()` y nos permite filtrar un conjunto de datos con una característica en específico. Los ejemplos mostrados en este artículo se basarán en información de los data set *iris* y *mtcars*. Antes de comenzar, haremos un resumen de los elementos que contiene cada data set.

```
str(iris)
summary(iris)
glimpse(iris)

str(mtcars)
summary(mtcars)
glimpse(mtcars)
```

Nota_ `glimpse()` es una función de dplyr que permite observar un resumen de la información presente en el data frame de trabajo.