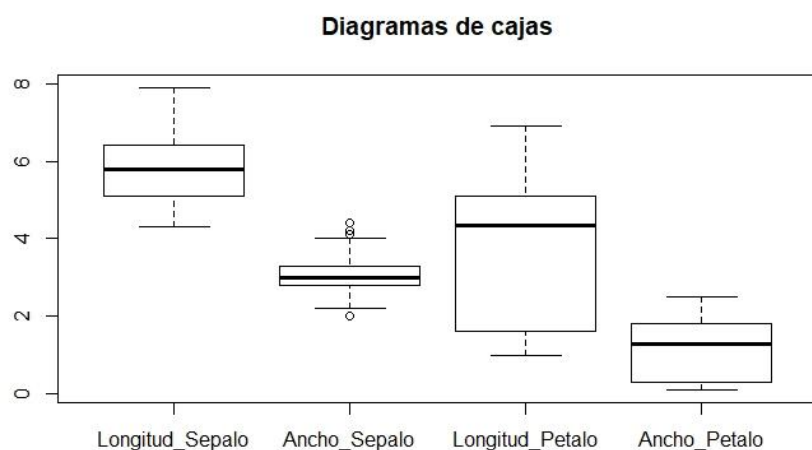


Comentario Data Set Iris

La base de datos Iris está compuesta por 150 casos y 5 variables. En el gráfico 1 se puede ver un boxplot de las variables numéricas de la misma.

Gráfico 1:



Se puede observar que hay unas variables con un mayor rango que otras, siendo el mayor caso el de la longitud del pétalo, y el menor el del ancho del pétalo.

A continuación, en el gráfico 2 se observa la relación entre las cuatro variables a través de la correlación que existe entre ellas.

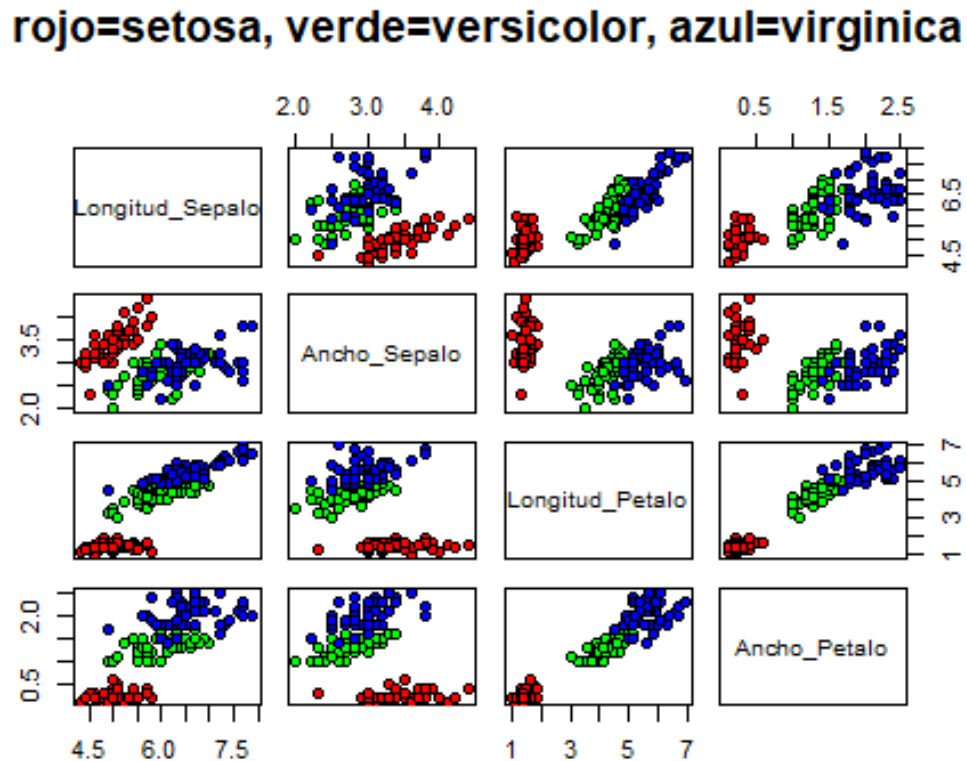
Se puede ver que la longitud del sépalos tiene una correlación de 0.8 y 0.9 con la longitud del pétalo y el ancho del pétalo respectivamente. Por otro lado, la longitud y el ancho del pétalo tienen una correlación de 1, es decir, la máxima posible. Mientras el ancho del sépalos tiene unas correlaciones más bajas con el resto de las variables.

Gráfico 2:



Por último, en el gráfico 3 se muestra la relación que guardan las variables entre sí, pero a través de una nube de puntos, donde el color de los puntos es la especie de flor a la que pertenece cada caso. Si nos fijamos en las relaciones señaladas anteriormente en la correlación, vemos que se producen dos divisiones claras, por un lado la especie setosa (color rojo), y por otro una mayor agrupación de las especies versicolor (verde) y virginica (azul). Por lo que se ve que hay una división en las características de la especie setosa frente a las otras dos.

Gráfico 3: Elaboración propia a partir de los datos de Anderson y Fisher (Wikimedia, 2019).



Bibliografía:

Anderson, E., & Fisher, R. (18 de octubre de 2019). *Wikimedia*. Obtenido de Wikimedia: https://commons.wikimedia.org/wiki/File:Iris_dataset_scatterplot.svg

El código del informe se puede encontrar en el siguiente repositorio:

https://github.com/miguellgpm/MasterDataScienceCUNEF/tree/master/Dimension_Reduction/Practica_1_IRIS