UNIVERSIDAD CATOLICA BOLIVIANA “SAN PABLO”

MAESTRIA EN CIENCIA DE DATOS, TERCERA VERSION



Materia: ANALISIS ESTADÍSTICO II

Practica No.4

Maestrante: Ramón Wilder Serdán Cárdenas

Julio 2022

La Paz – Bolivia

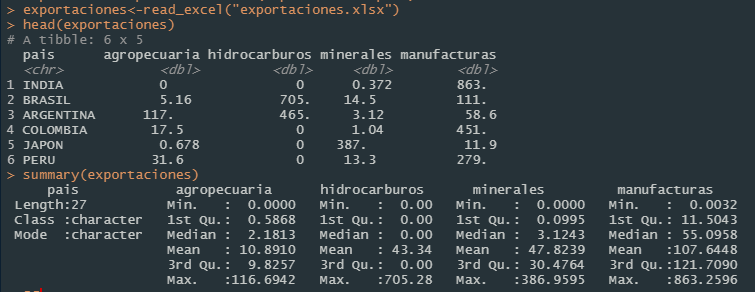
Contenido

[1.- Realizar un ejercicio de Árboles de Decisión 3](#_Toc108765940)

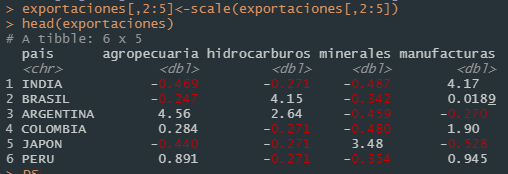
[2.- Realizar un ejercicio de Análisis de Componentes Principales. 7](#_Toc108765941)

## 1.- Realizar un ejercicio de Clustering

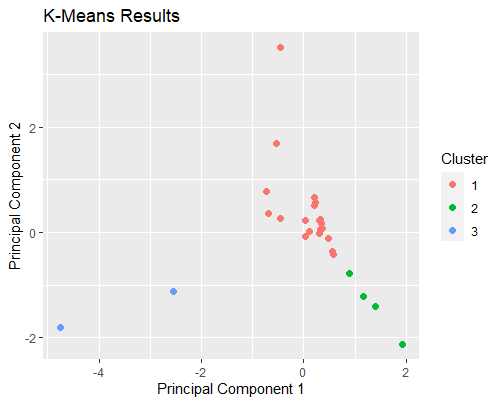
Mediante *clustering* analizar los datos de exportaciones acumuladas a mayo de 2022 por país destino y agrupación TNT (tradicional no tradicional) publicada por el INE y expresado en millones de dólares.



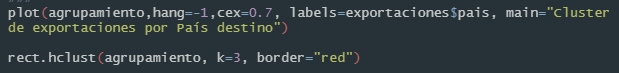
Realizamos una transformación en la escala de las observaciones:

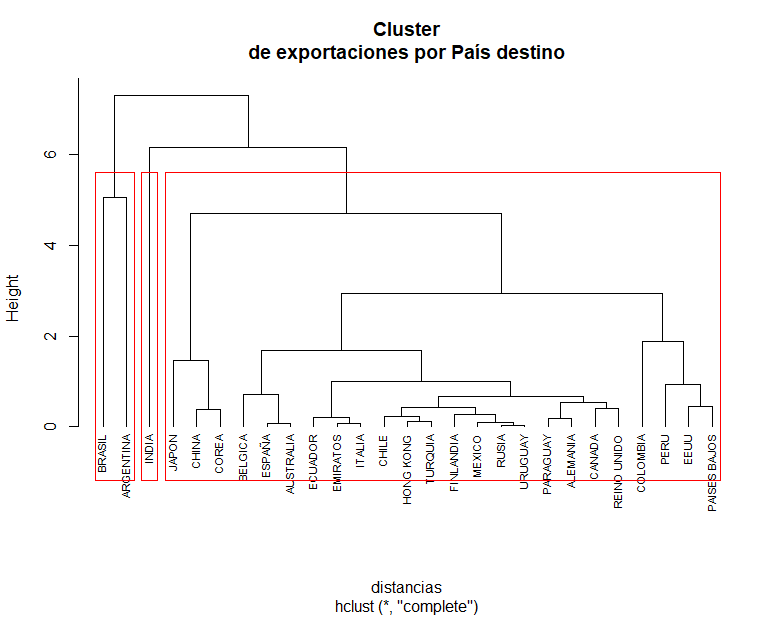


Como una primera aproximación para la realización del cluster jerárquico, utilizamos un algoritmo de k-means con 3 centros para visualizar los grupos en los datos.



Ahora, utilizamos la función de plot para mostrar los Clusters formados mediante código.



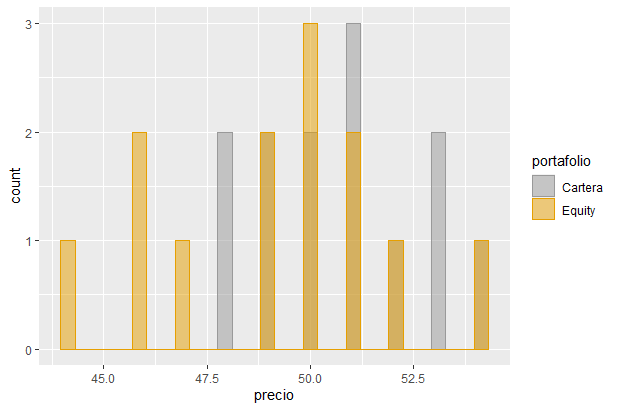
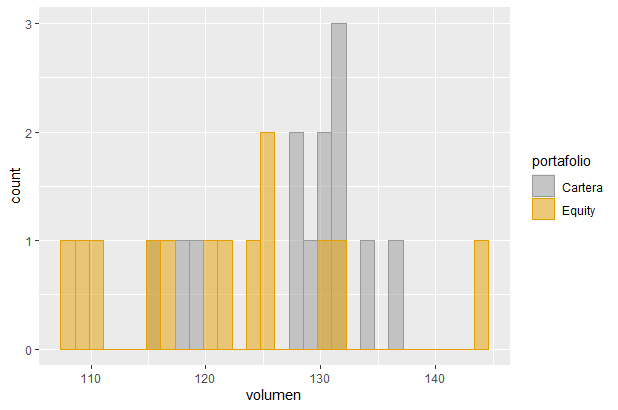
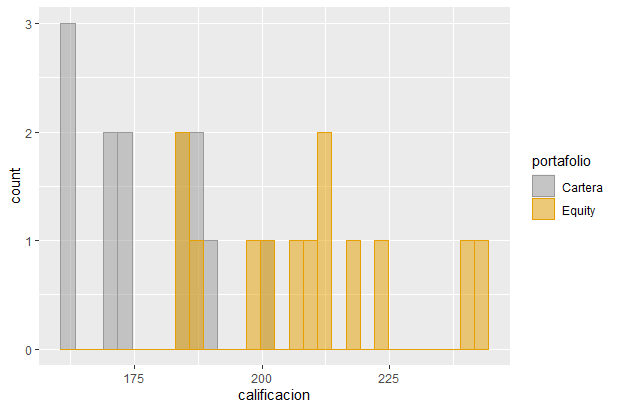


El Clustering realizado por código muestra grupos coherentes con la tendencia de exportaciones de Bolivia. Los principales socios comerciales son Brasil y Argentina por venta de hidrocarburos e India por la exportación de minerales. Los demás países son destinos de exportaciones varias como manufacturas, productos agropecuarios

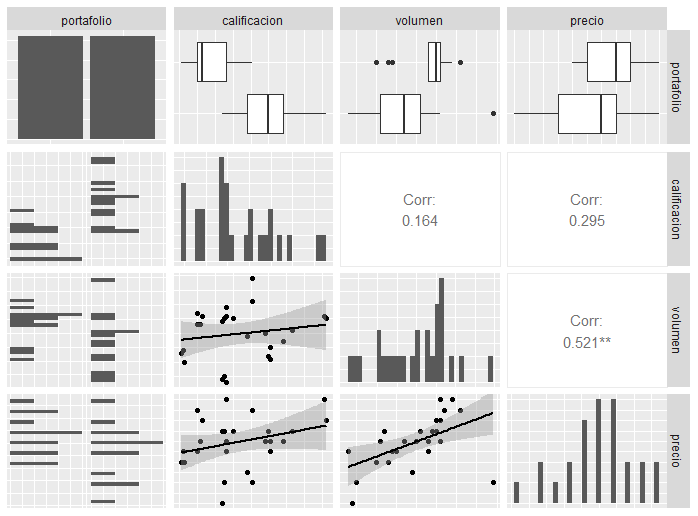
## 2.- Realizar un ejercicio de Análisis discriminante.

En un portafolio de 2 tipo de instrumentos, se cuenta con información de 13 instrumentos de Cartera y 13 de equity. Se requiere si un nuevo instrumento pertenece a algún tipo de portafolio.

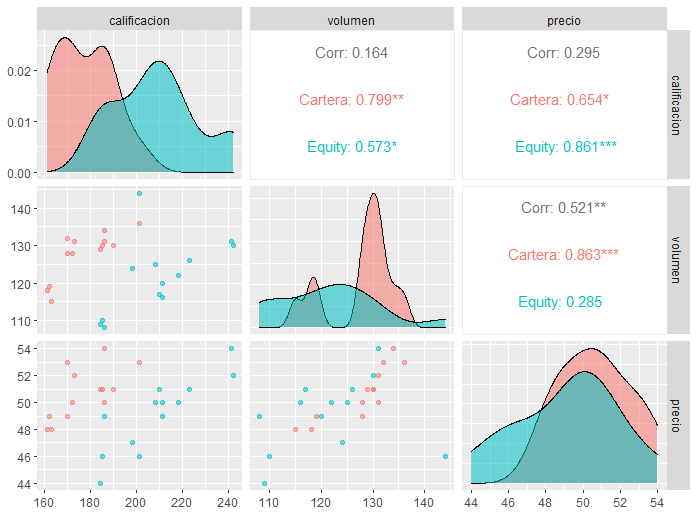
Realizamos el análisis de las 3 características de cada instrumento:



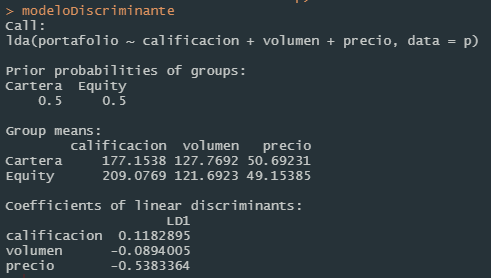
Se observa una alta correlación entre el precio del instrumento y el volumen.



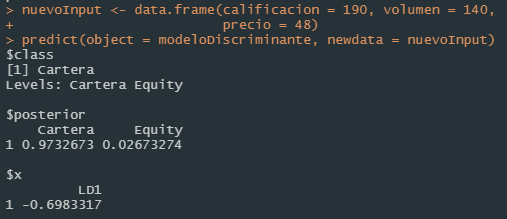
La distribución de la variable precio es comporta de manera normal, en contraposición de las variables volumen y calificación.



El modelo discriminantes sería el siguiente:



La predicción para una nueva entrada al dataset es la siguiente:



Con una alta probabilidad de pertenecer al tipo de portafolio Cartera. Asimismo, observamos en el grafico el resultado de la partición mediante análisis discriminante.

