**Instituto Politécnico Nacional**

**Escuela Superior de Cómputo**

**Programación para Ciencia de Datos**

**Práctica 8, parcial 2**

**De Luna Ocampo Yanina**

**Galindo Durán Cristal Karina**

**3MA1**

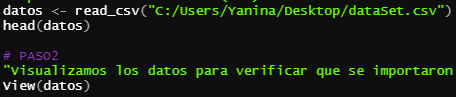
**Fecha: 23/11/2021**

***Ejercicio1.***

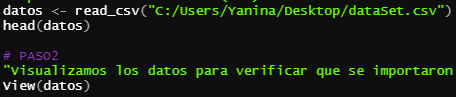
1. Descargar el archivo de ventas clientes del siguiente enlace.
2. Aplica el Análisis de Componentes Principales.
3. Realiza la interpretación de los resultados.

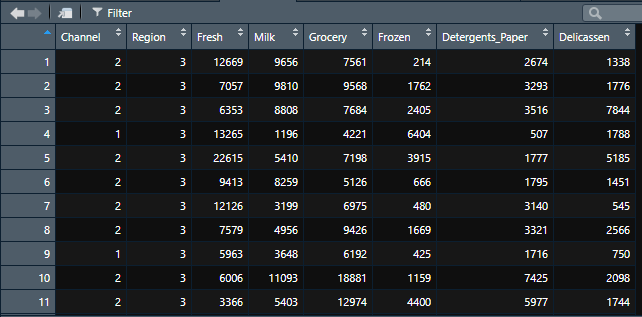
***Procedimiento:***

Descargamos los datos del link dado, lo importamos a R con las funciones vistas en clase.

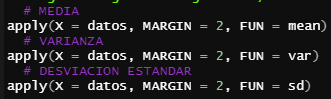


y procedemos a visualizarlos para ver que se hayan importado de la forma correcta.

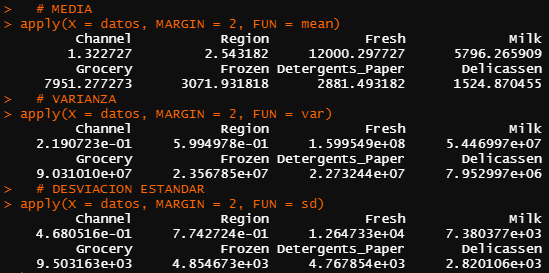




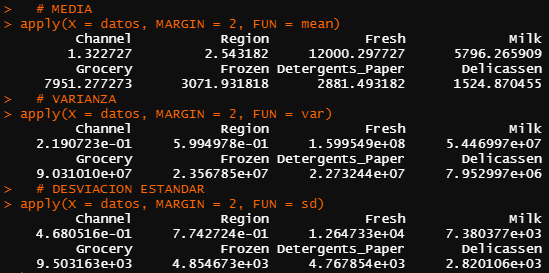
Obtenemos la media, la varianza y la desviación estándar con el fin de ver que no sean un cero y poder aplicar el método pedido correspondiente, ya que, si alguna llega a ser igual a 0, no tiene sentido aplicar el método a desarrollar.



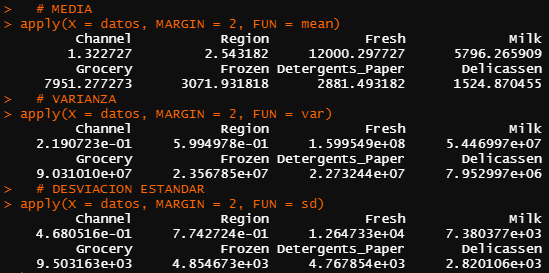
La media menor pertenece a Channel y más alta pertenece a la categoría Fresh. Podemos observar que Fresh es la 12000 más grande de lo que se obtuvo en Channel. Recordamos que para sacar la media se utiliza la función “mean”.



La varianza con una magnitud superior al resto es la de Grocery y la de menor magnitud pertenece a Fresh, contrario a su media. Recordamos que para sacar la varianza se utiliza la función “var”.



La desviación estándar con una magnitud superior al resto es la de Grocery y la de menor magnitud pertenece a Fresh, contrario a su media. Recordamos que para sacar la desviación estándar se utiliza la función “sd”.

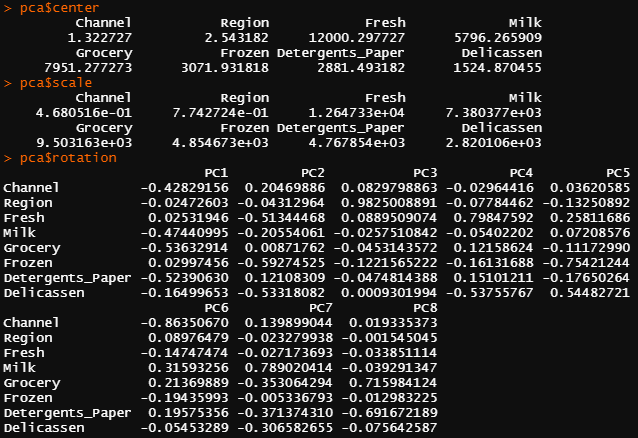


La función siguiente [prcomp()] es una de las que podemos utilizar en R para poder obtener nuestro PCA.

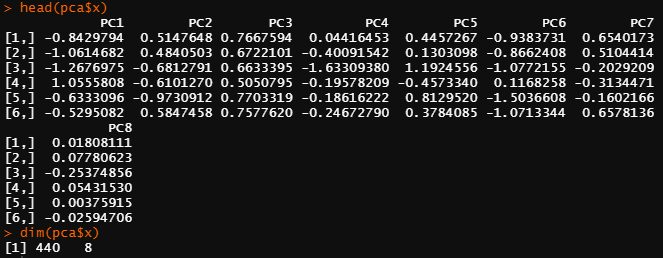
Por defecto centran las variables para que tengan media cero. Utilizamos el prcomp como dijimos previamente para poder empezar a implementar nuestro modelo estadistico que nos ayudará a analizar los datos datos.

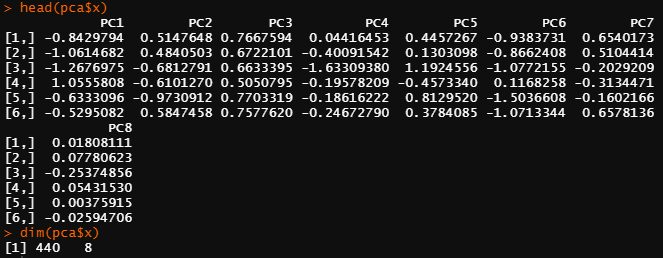


Mandamos llamar individualmente los subgrupos de arriba. Con center y scale obtenemos la media y la desviación típica de las variables, en cuanto al rotation, podemos ver que analiza el vaor de los loading de cada eigenvector.

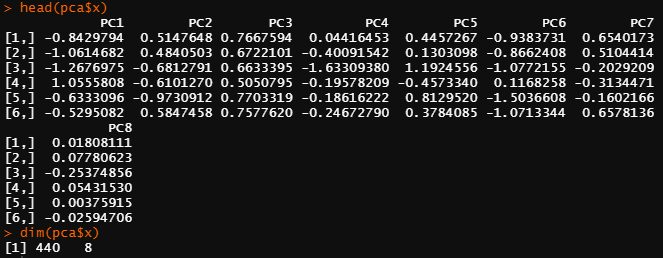


Asimismo, calcula autimáticamente el valor de los componentes principales para cada observación, multiplicando los datos por los vectores de loadings. El resultado se almacena en la matriz x.





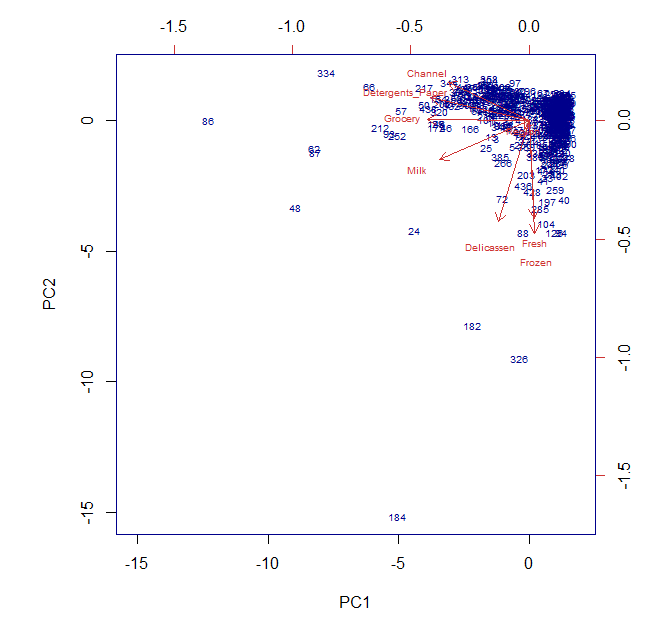
Obtenemos dim de lo que acabamos de sacar.



***Resultado:***

Obtenemos como gráfica final la siguiente:





***Ejercicio2.***

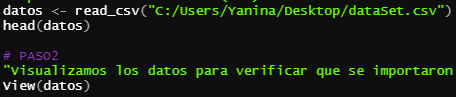
1. Descargar el archivo de tu interés.
2. Aplica el Análisis de Componentes Principales.
3. Realiza la interpretación de los resultados.

***Procedimiento:***

Descargamos los datos del dataset de nuestro interés, lo importamos a R con las funciones vistas en clase.

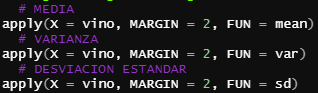


y procedemos a visualizarlos para ver que se hayan importado de la forma correcta.

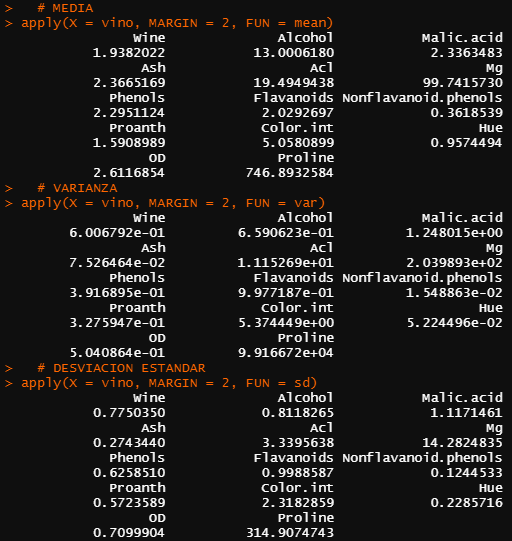




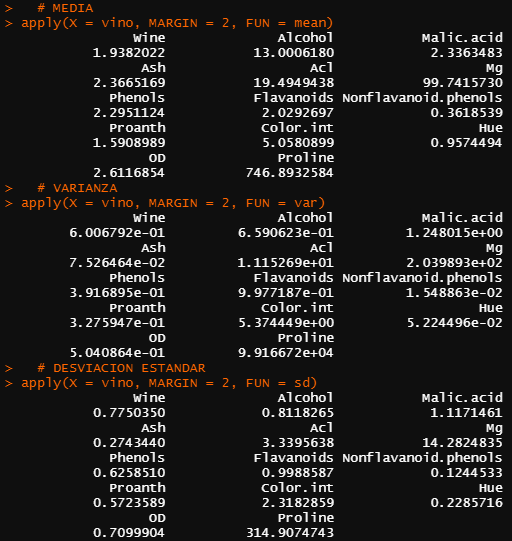
Obtenemos la media, la varianza y la desviación estándar con el fin de ver que no sean un cero y poder aplicar el método pedido correspondiente, ya que, si alguna llega a ser igual a 0, no tiene sentido aplicar el método a desarrollar.



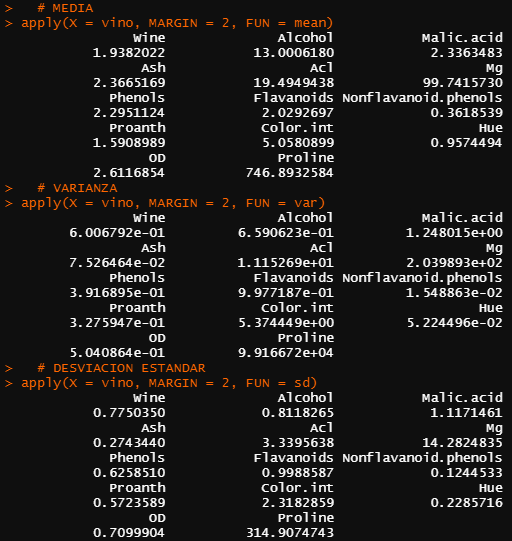
La media menor pertenece a Nonflavanoid.phenols y más alta pertenece a la categoría Proline. Podemos observar que Proline es al menos 2,068 veces más grande de lo que se obtuvo en Nonflavanoid.phenols. Recordamos que para sacar la media se utiliza la función “mean”.



La varianza con una magnitud superior al resto es la de Flavanoids y la de menor magnitud pertenece a Acl. Recordamos que para sacar la varianza se utiliza la función “var”.



La desviación estándar con una magnitud superior al resto es la de Proline y la de menor magnitud pertenece a Nonflavanoid.phenols, al igual que su media. Recordamos que para sacar la desviación estándar se utiliza la función “sd”.



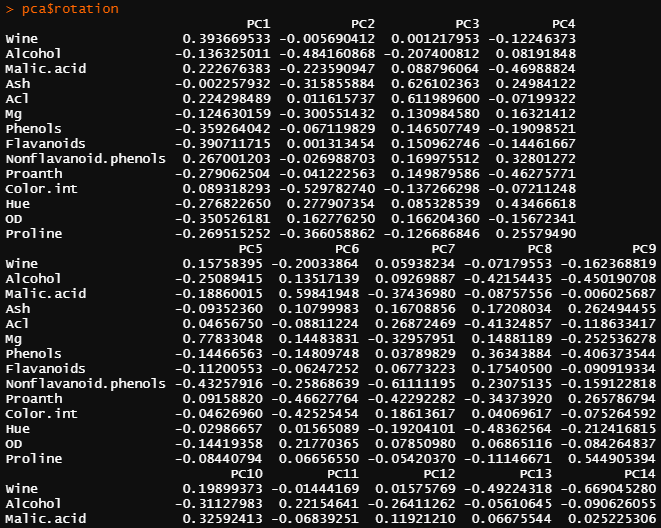
La función siguiente [prcomp()] es una de las que podemos utilizar en R para poder obtener nuestro PCA.

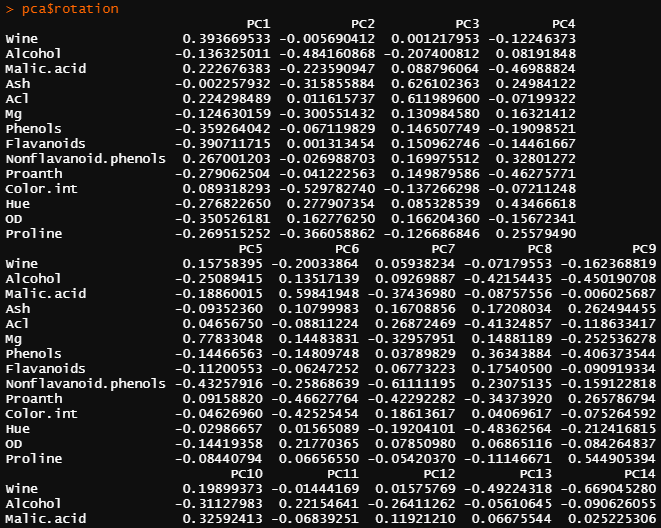
Por defecto centran las variables para que tengan media cero. Utilizamos el prcomp como dijimos previamente para poder empezar a implementar nuestro modelo estadistico que nos ayudará a analizar los datos datos.

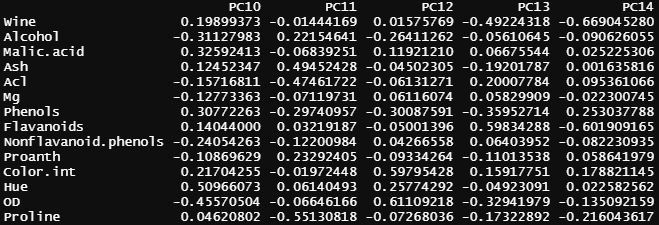


Mandamos llamar individualmente los subgrupos de arriba. Con center y scale obtenemos la media y la desviación típica de las variables, en cuanto al rotation, podemos ver que analiza el vaor de los loading de cada eigenvector.

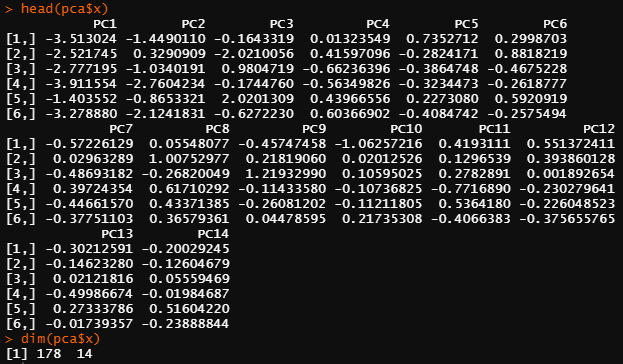




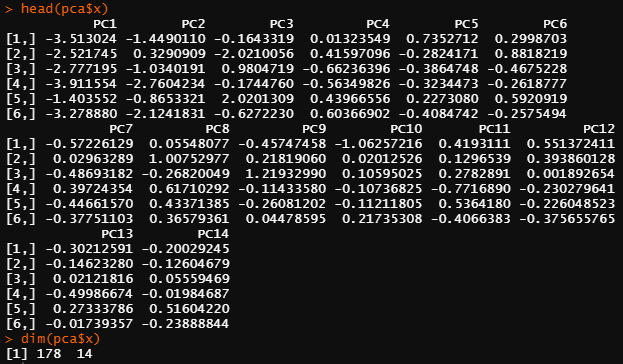




Asimismo, calcula autimáticamente el valor de los componentes principales para cada observación, multiplicando los datos por los vectores de loadings. El resultado se almacena en la matriz x.



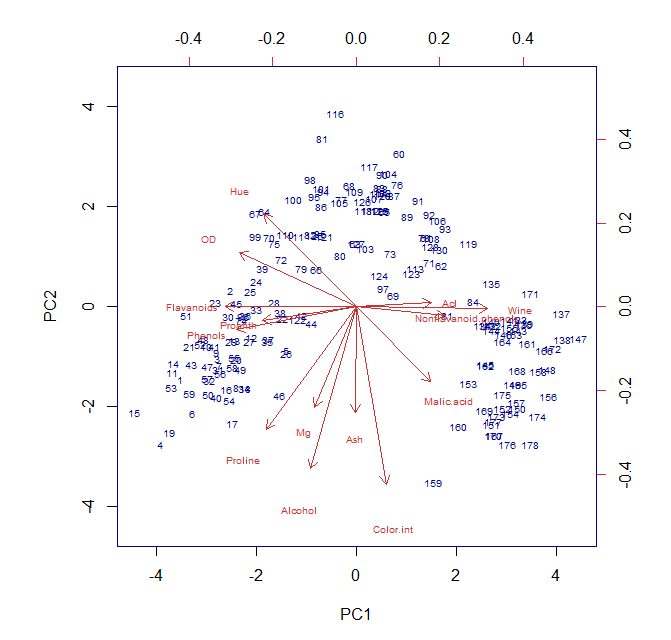
Obtenemos dim de lo que acabamos de sacar.



***Resultado:***

Obtenemos como gráfica final la siguiente:





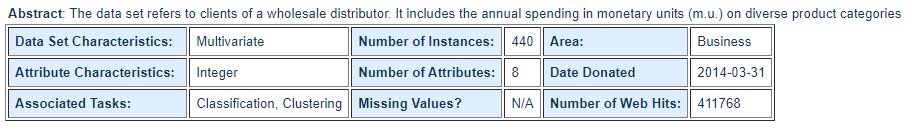
¿Qué aprendí con esta práctica?

Reforcé la importación de datasets en R. Aprendí a implementar un nuevo método estadístico que nos ayuda a simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez que conserva su información.

Nos permite condensar la información apartada por múltiples variables en pocos componentes. Esto lo hace un método muy útil de aplicar previa utilización de otras técnicas estadísticas.

***Información de los DataSets:***

***Data1, ventas:***





***Data2, wine:***

***Fuente:*** <https://archive.ics.uci.edu/ml/datasets/wine>

