

## BIPLOT

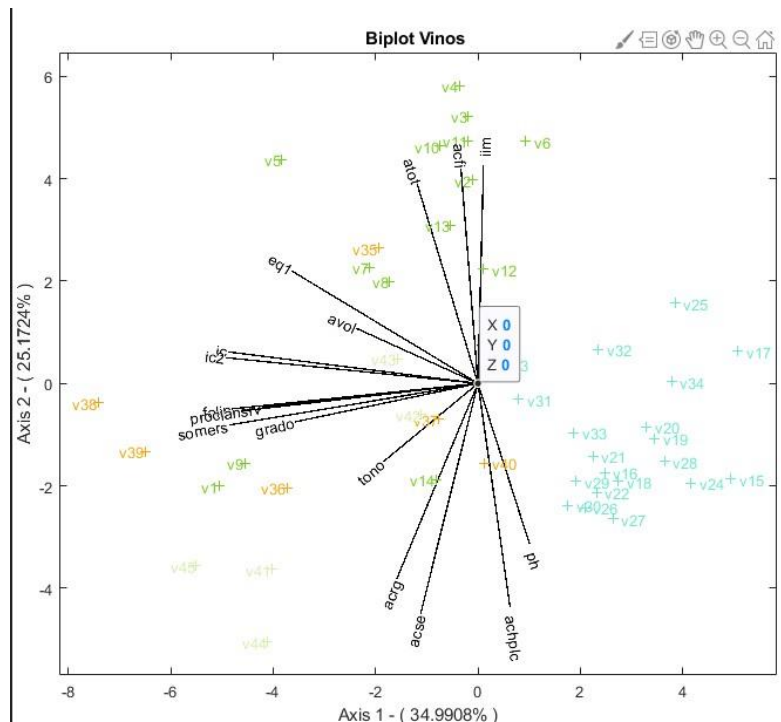
Para el ejemplo se utiliza la base de datos “Vinos”, contiene datos de los distintos factores presentes de 4 marcas de vinos.

Variables		
a_o	ph	achpic
denomina	folin	ic
grupo	somers	ic2
grado	siv	tono
avol	procian	iim
atot	acrg	eq1
acfi	acse	vla

Un biplots es un gráfico que representa la generalización multivariante de un diagrama de dispersión mediante variables. Los diagramas de dispersión muestran como las variables se distribuyen en un biplot, además, puede representar 3 o mas variables.

Un biplot es la aproximación de la distancia sobre alguna muestra multivariante en un espacio de dimensión reducida, sin embargo, se emplea de dimensión 2, para medir las variables de la muestra se realiza una superposición.

### Gráfico de variables utilizadas



Se observa que la variable “vla” no se encuentra bien representada, ya que su línea de proyección es muy corta. Se observa que las líneas que son largas tienen muy buena representación, mientras que mas corta sea la línea tiene una mala representación.

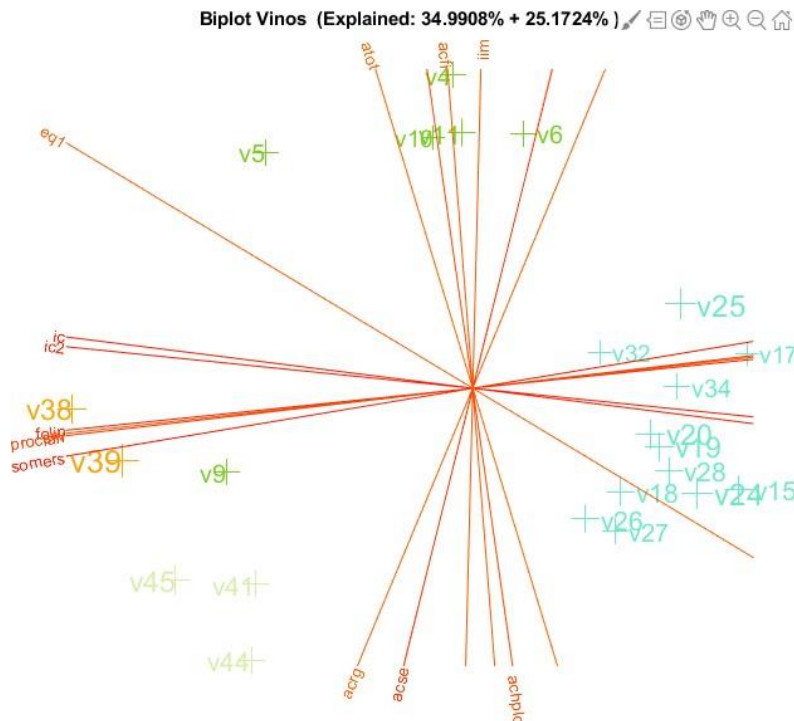
En este biplot las variables que no tienen buena representación son: tono, avol, grado y ph.

Variables con buena representación son: iim, acfi, atot, eq1, ic, ic2, somers, acrg, acse y achplc.

Al observar que algunas líneas de las variables estadísticamente son casi similares y es posible descartar algunas.

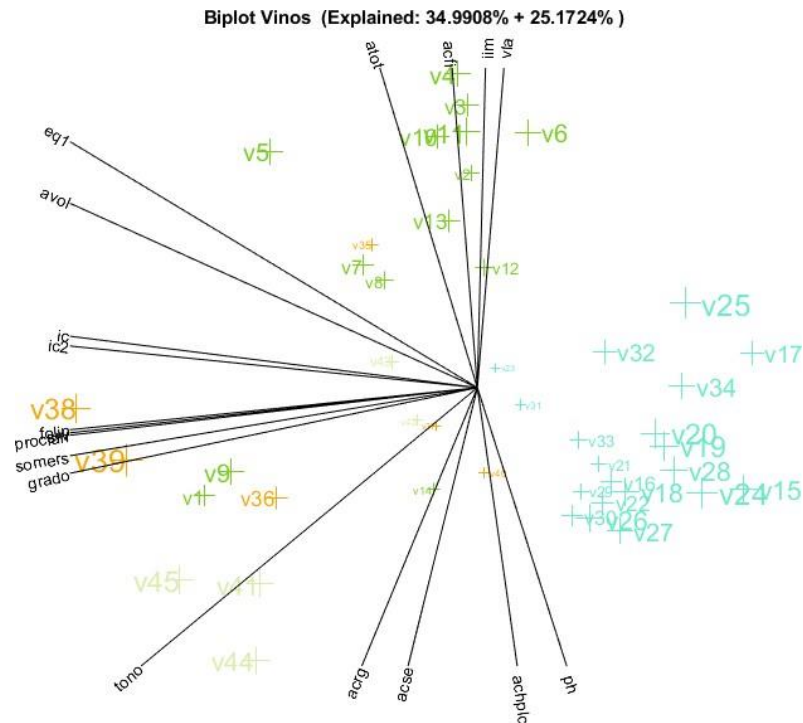
### Agrupaciones (Cluster)

La distribución de los datos es de la siguiente manera:



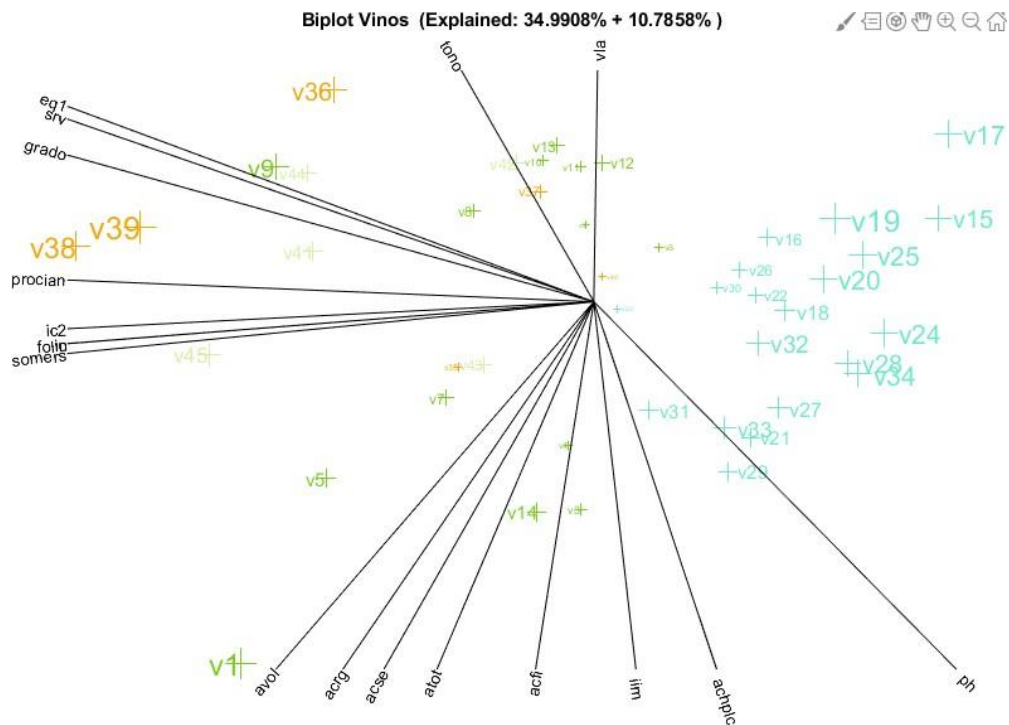
Se hizo un filtrado y se observa que las líneas gruesas representan una mejor calidad en las variables como “procian”.

Se realizan proyecciones con diferentes componentes en el siguiente gráfico:



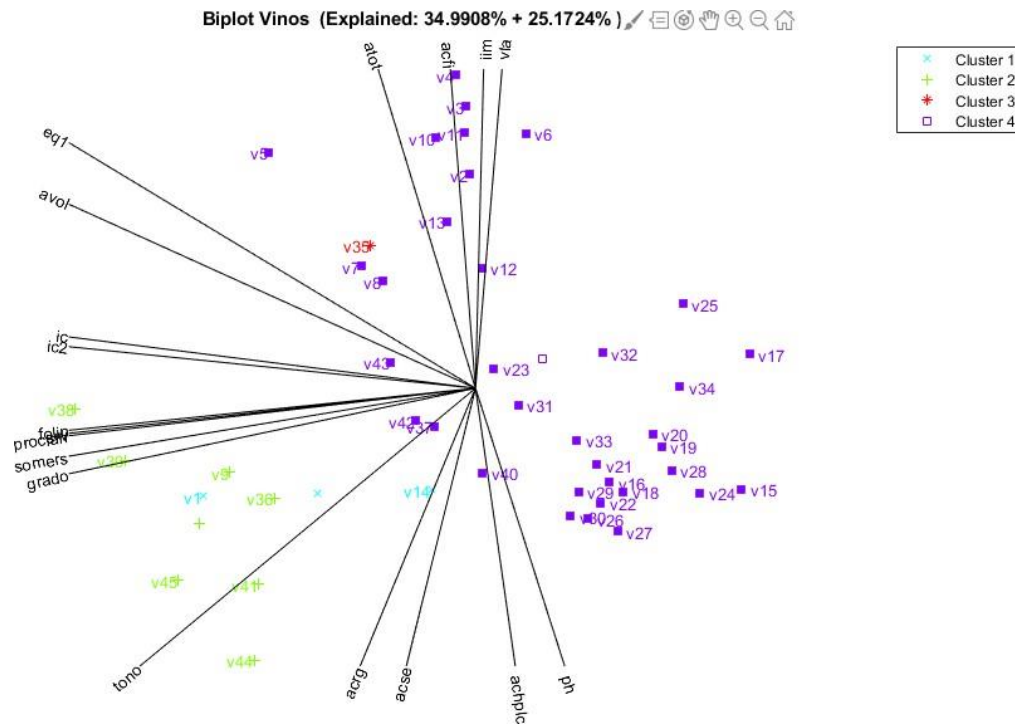
Se observa las proyecciones de los componentes 1 y 2, representan un 35% y 25%, por lo tanto, tienen un 60% de varianza explicada.

### Proyección de los componentes 1 y 3



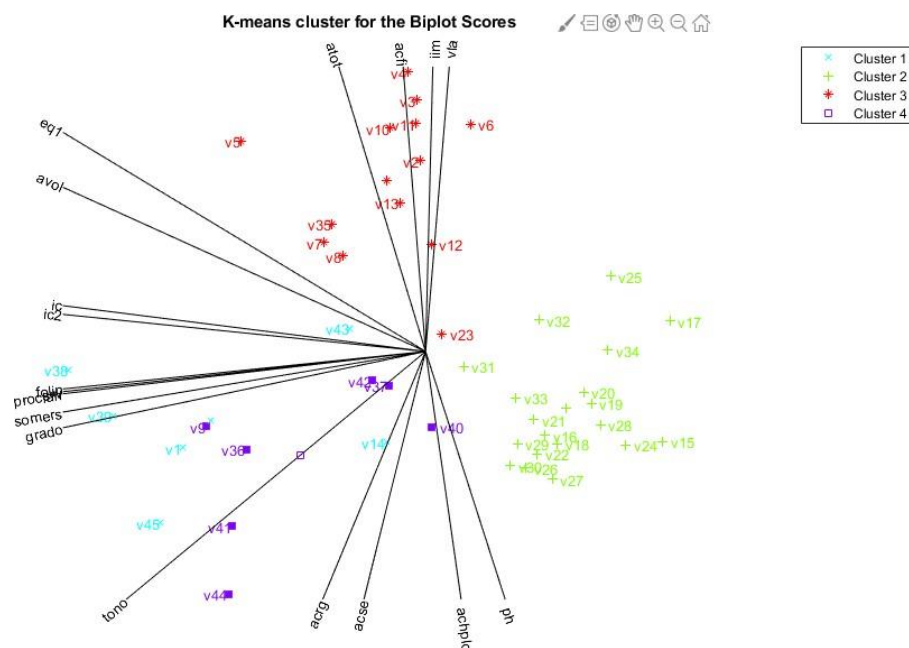
Estas proyecciones tienen mayor varianza, dado que el componente 1 tiene un 35% y el componente 2 tiene un 11% en total tienen un 46% de varianza explicada.

### Técnica de clúster para agrupaciones



Se observa los agrupamientos con características similares

### Gráfico de *K medias*



Se observa los datos mejor acomodados, ya que los datos se encuentran acomodados dentro de sus grupos.

# Cambio de gráficos para resaltar los grupos

