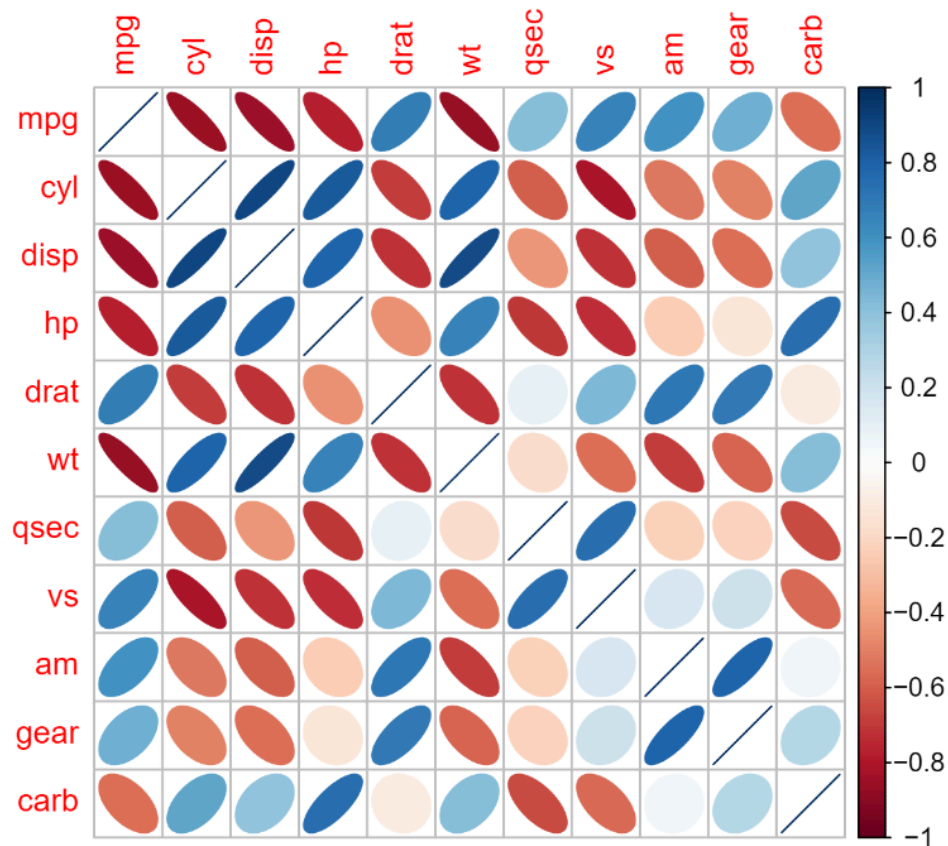


Actividad Evaluable 5: Análisis de los datos

Análisis de correlación



Las variables con correlación que podemos ver es la cilindrada de un vehículo con el consumo de galones por milla, esto hace sentido ya que mientras más cilindros mayor combustión.

La correlación del desplazamiento con la cilindrada también es muy grande debido a que motores que desplazan mayor volumen tienen mayor cantidad de cilindros.

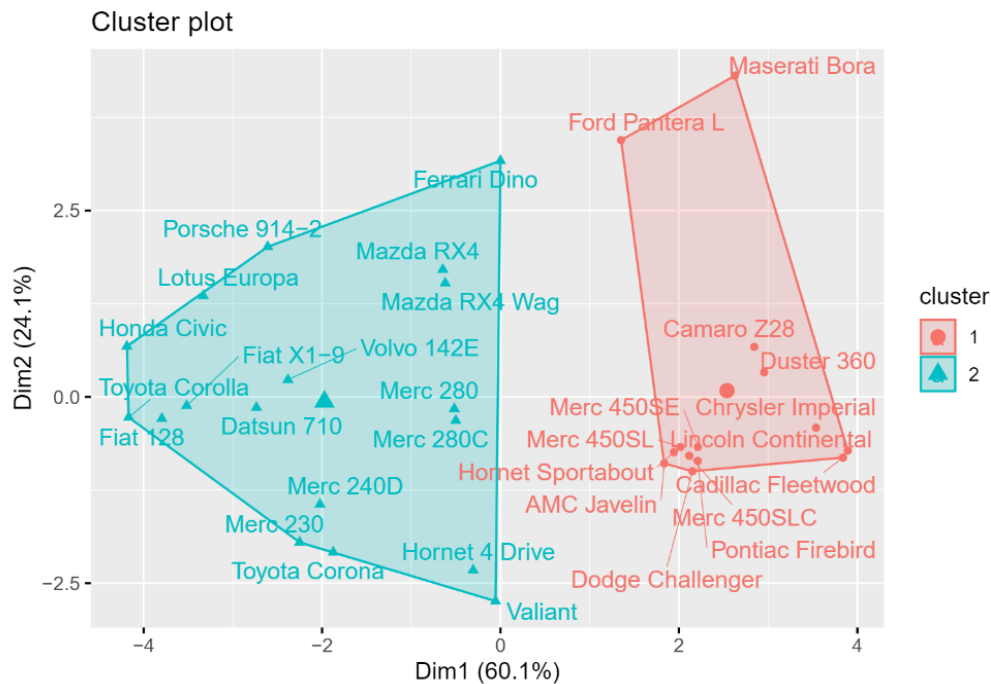
La correlación entre los caballos de fuerza y los cilindros también es notoria ya que la idea de un motor más grande es mayor potencia.

También podemos ver que los motores en v tendían a tener transmisiones manuales así como un mejor desempeño en el cuarto de milla.

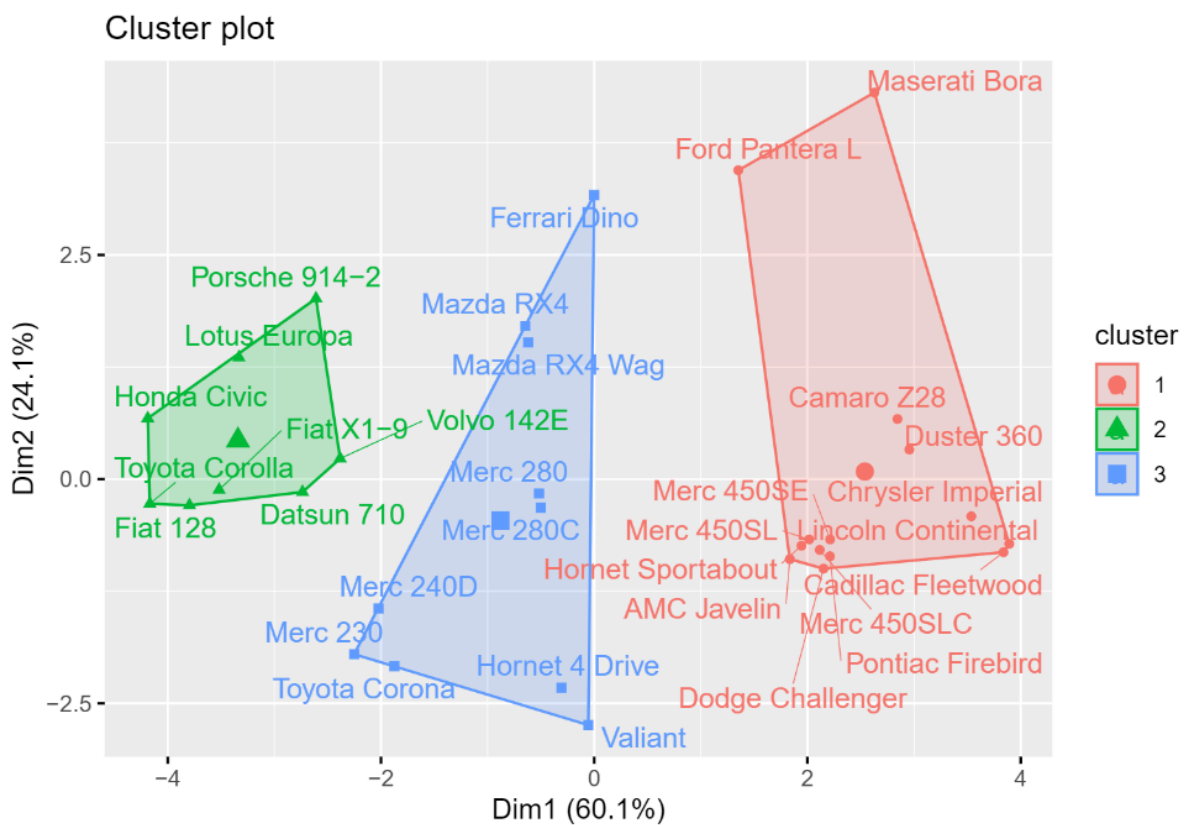
Los coches manuales tienen una mayor cantidad de cambios en la caja.

Por último, la correlación entre el peso, y la cilindrada, tamaño del motor así como el consumo, aumentan a menor peso.

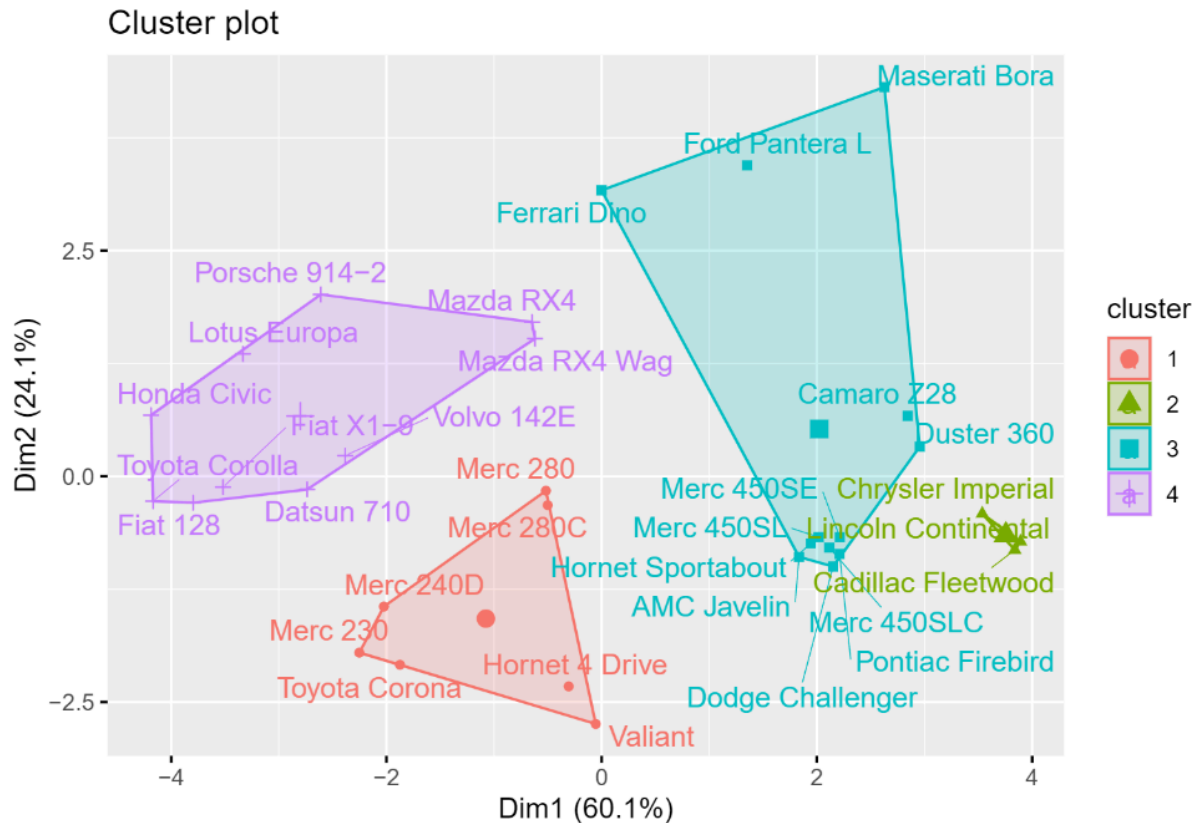
Análisis de K-means



Podemos observar que al realizar el método de agrupamiento k-means con 2 Clusters, tenemos una separación muy amplia entre los Ford Pantera L y el Maserati Bora, ya que estos autos tienen un número diferente de carburadores. Los clusters los divide por peso y por consiguiente el tamaño del pistón y la cilindrada.



Podemos observar que al realizar el método de agrupamiento k-means con 3 Clusters, que igualmente que la agrupación con 2 clusters, los segmenta por peso y por lo tanto tamaño general del motor y después por los carbs.



Podemos observar que al realizar el método de agrupamiento k-means con 4 Clusters, el grupo verde es un segmento de camionetas, los más pesados del dataset, esto se cumple para los otros grupos, después toma los cambios en caja y ya por ultimo toma en cuenta los carburadores.

Conclusión

Tenemos un dataset con variables que tienen un alto valor de correlación, el análisis nos muestra hechos.

El mejor número de clusters para del k means depende de que necesitas, por ejemplo si buscamos un coche con potencia utilizamos el de 3 clusters, en otro caso si quisiéramos evaluar una caja de cambios con varias velocidades podemos utilizar la agrupación de 4 clusters.

Entonces podemos decir que si tenemos una dataset con altos valores de correlación y buscamos aplicar un método de agrupación K-Mean, el mejor número de clusters dependerá de lo que busquemos con nuestro dataset.