

Data Set Freedman

Las variables de este dataset Freedman son la población, no-blanco, densidad y crimen de diversas ciudades.

La variable población nos da información sobre la cantidad de habitantes en una ciudad (tenemos 100 observaciones), no-blanco sobre la cantidad de habitantes que no son blancos (tenemos 110 observaciones). Densidad nos da la densidad de población en una ciudad (tenemos 100 observaciones). Crimen nos da una estadística de delitos cometidos en cada ciudad (tenemos 110 observaciones).

hc2\$ac:

El coeficiente de aglomeración de un vértice en un grafo cuantifica qué tanto está de agrupado (o interconectado) con sus vecinos. Se puede decir que si el vértice está agrupado como un clique (grafo completo) su valor es máximo, mientras que un valor pequeño indica un vértice poco agrupado en la red. Para nuestros clusters nosotros obtuvimos un coeficiente de 0.93, esto quiere decir que el "grado" de interconexión entre los elementos es muy alto.

Dendograma hc3 ¡- agnes(data, method = "ward")

En este dendrograma estamos utilizando la función agnes, la cual utiliza la agrupación jerárquica aglomerativa y además obtiene el coeficiente de aglomeración, y el método ward, el cual minimiza la varianza total dentro del grupo. En cada paso, se fusionan el par de clústeres con una distancia mínima entre los clústeres. Ward forma grupos minimizando la pérdida asociada con cada grupo. En cada paso, se considera la unión de cada par de clústeres posible y se combinan los dos clusters cuya fusión da como resultado un aumento mínimo en la pérdida de información. Ward define la pérdida de información en términos de un criterio de suma de cuadrados de error (ESS).