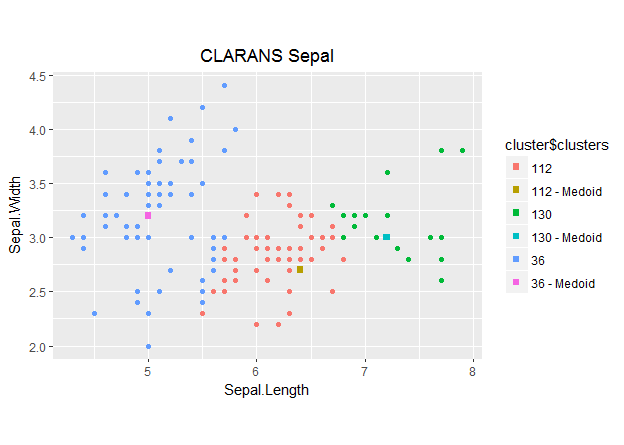
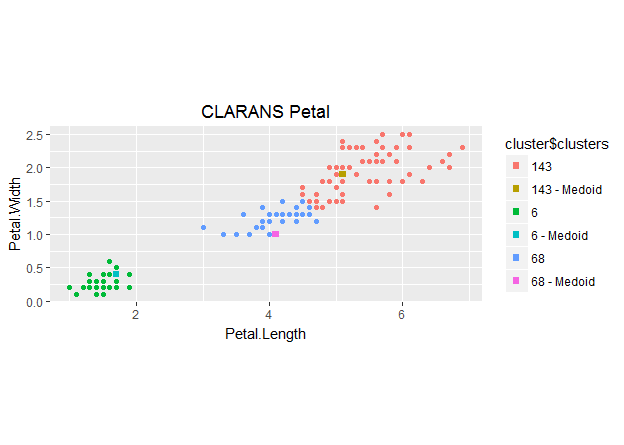
**MEMORIA PRÁCTICA CLUSTERING**

Resultados con CLARANS

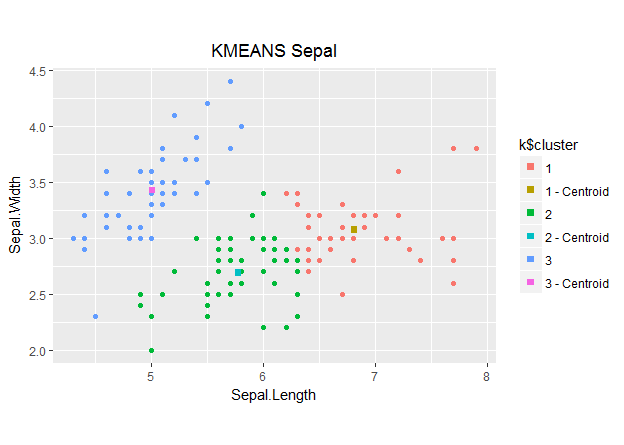


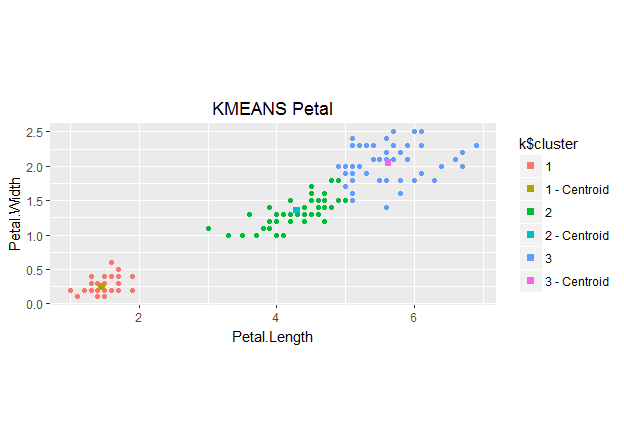
Error absoluto: 74.50096



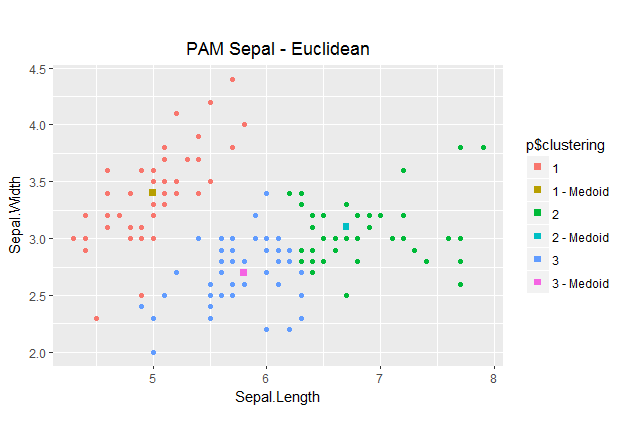
Error absoluto: 71.10767

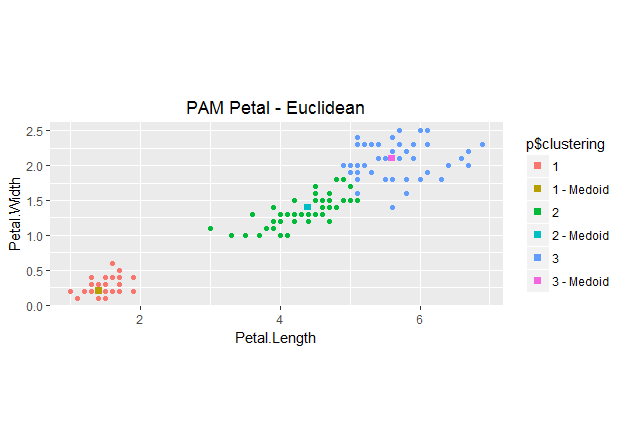
Resultados con KMEANS

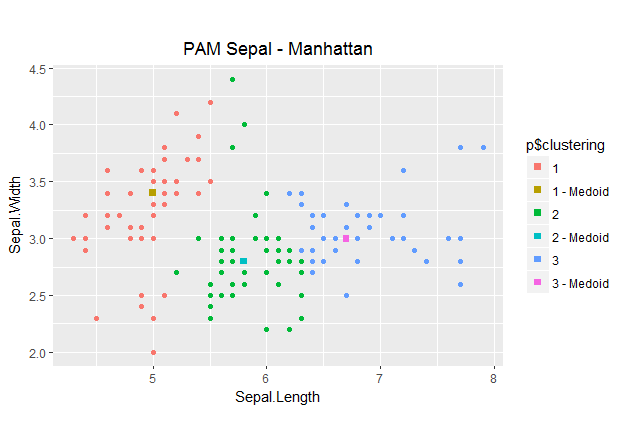


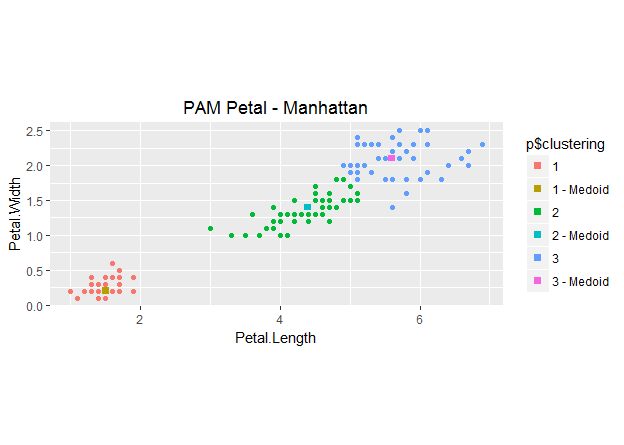


Resultados con PAM



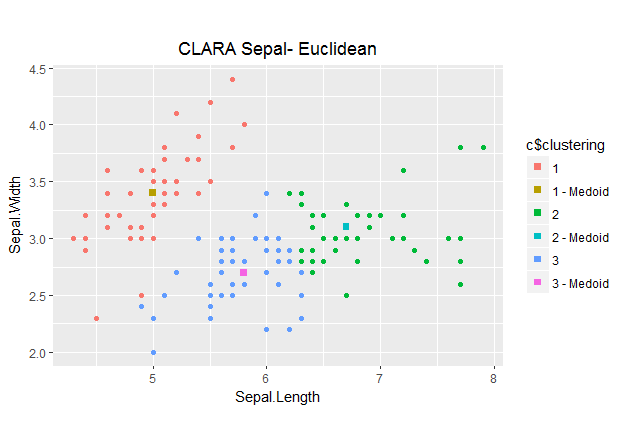


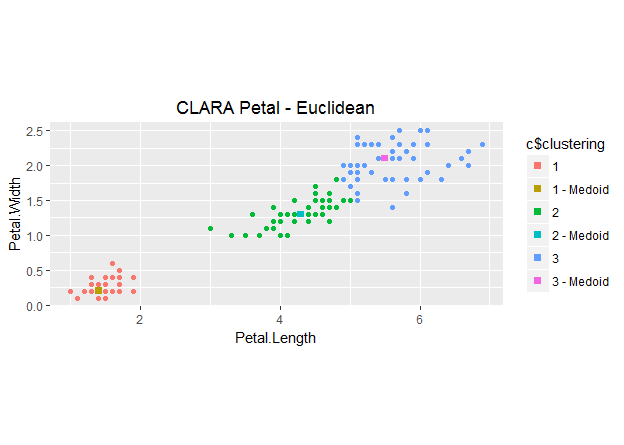


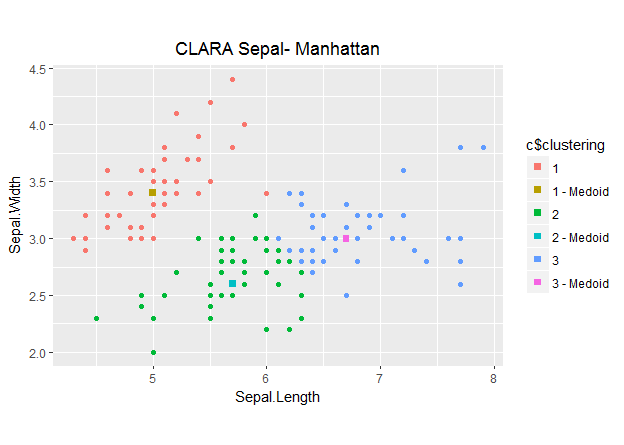


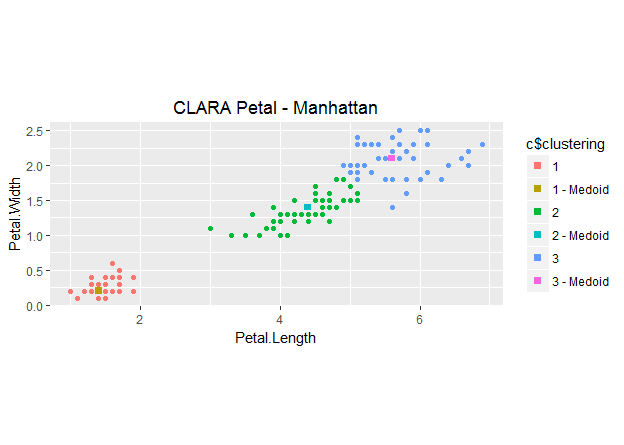
PAM Sepal Manhattan ha incluido valores en el cluster 1 (rojo) que quizá deberían estar en el cluster 2 (verde) y viceversa.

Resultados con CLARA









Tanto el algoritmo k-means como el PAM (algorito k-medoids) dividen el conjunto de datos en particiones y ambos intentan minimizar la distancia entre un punto y el centro seleccionado (maximizando la similitud entre los puntos de un mismo cluster). La diferencia está en el tipo de puntos que emplean como centros. Mientras que k-means trabaja con centros que son resultado de una media, los centros de PAM son los objetos más centrados en el clúster, lo que es la mediana (son datos que existen en el dataset).

PAM es menos sensible a los outliers comparado con k-means porque usa medianas y minimiza las disimilitudes entre clusters.

CLARA es una variación de PAM usada para datasets grandes.

**Observaciones en las gráficas:**

Cuando usamos la distancia Manhattan tanto en PAM como en CLARA los errores son mayores. Creemos que esto se podría deber a que la métrica sea menos apropiada al tratar con valores continuos.

El hecho que incluya outliers en clusters en los que no debería, podría deberse a que no son suficientes iteraciones para tratar de mejorar el error. Si cogemos un centro de un grupo, y al calcular el error tenemos en cuenta algún outlier, cuya distancia al centro es muy grande, aumentará mucho el resultado del cálculo, y por lo tanto, lo más probable es que no mejoremos el error previo y perderemos una iteración.

Si el dataset tiene outliers, puede afectar al resultado del clustering desplazando los centros.

En CLARANS, los centros de los clusters 6 (verde) y 68 (azul) están posicionados bastante al margen del grupo, en vez de centrados. Se nos ocurre que esto podría depender de las instancias que se hayan seleccionado aleatoriamente al principio y el número de veces que haya mejorado el error. Por ejemplo, si los centros aleatorios iniciales han “caído” en el margen de los grupos y, en el peor caso, no hemos mejorado ninguna vez, el algoritmo acaba con los centros en el margen, sin cambiar.