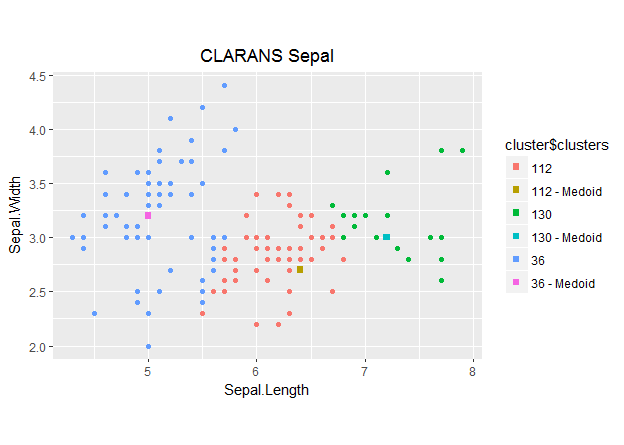
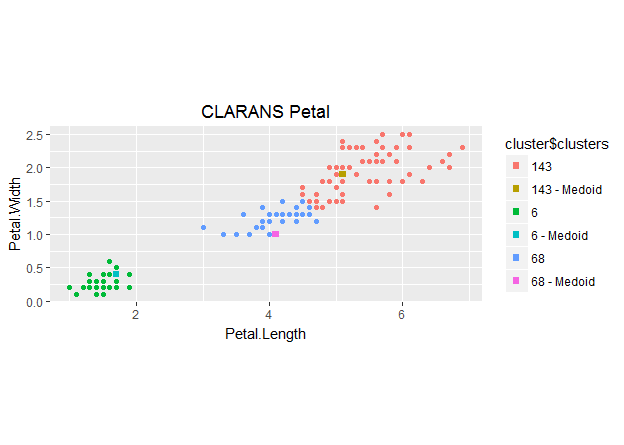
**MEMORIA PRÁCTICA CLUSTERING**

Resultados con CLARANS

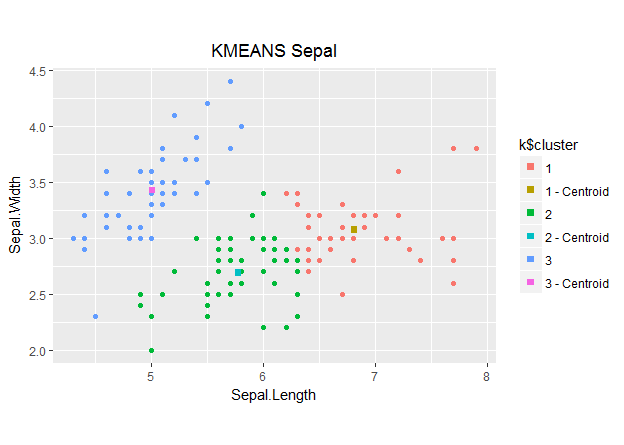


Error: 74.50096

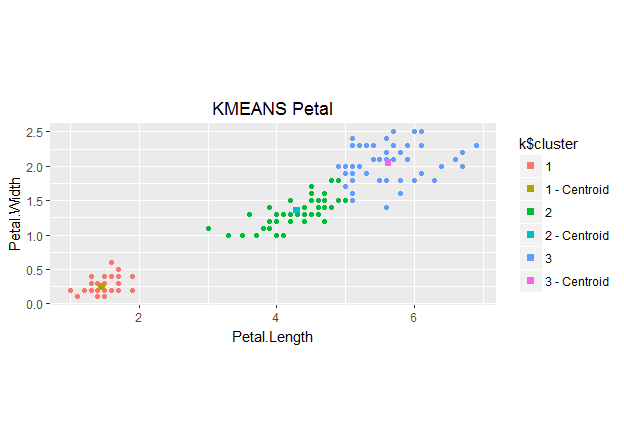


Error: 71.10767

Resultados con KMEANS

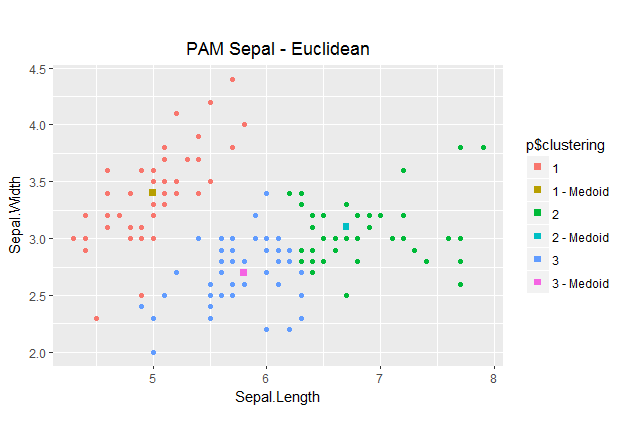


Error: 130.4753

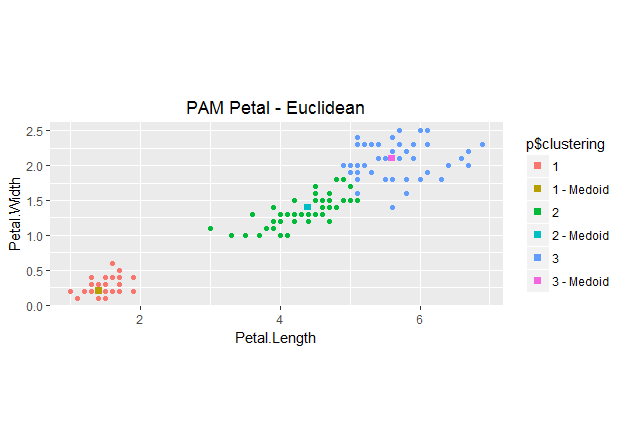


Error: 550.8953

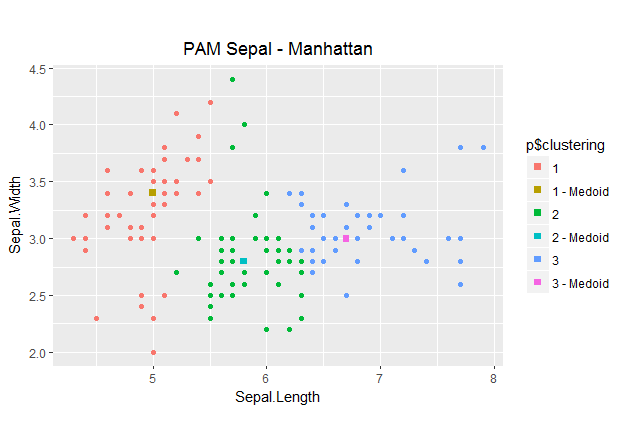
Resultados con PAM



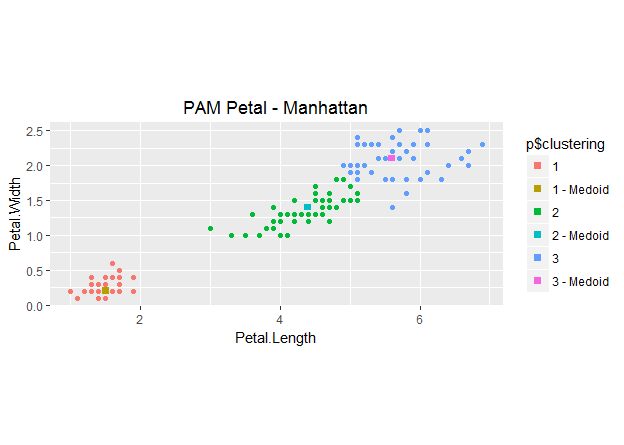
Error: 0.4144793



Error: 0.3663135



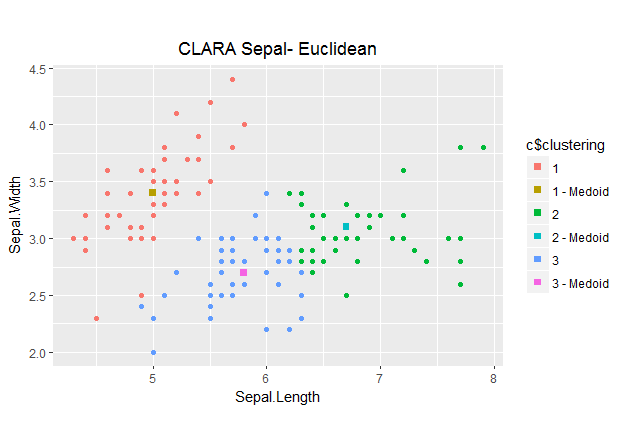
Error: 0.5306667



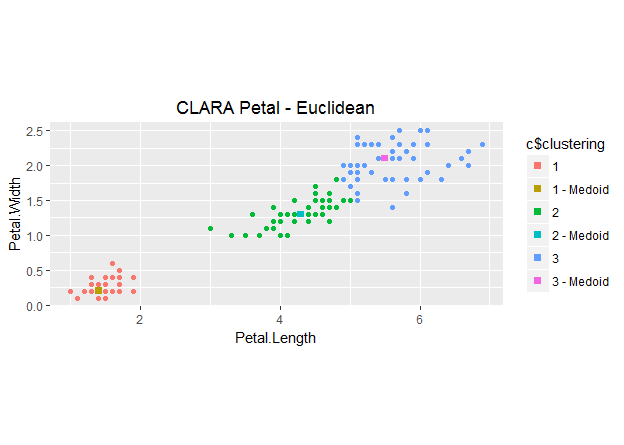
Error: 0.4526667

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

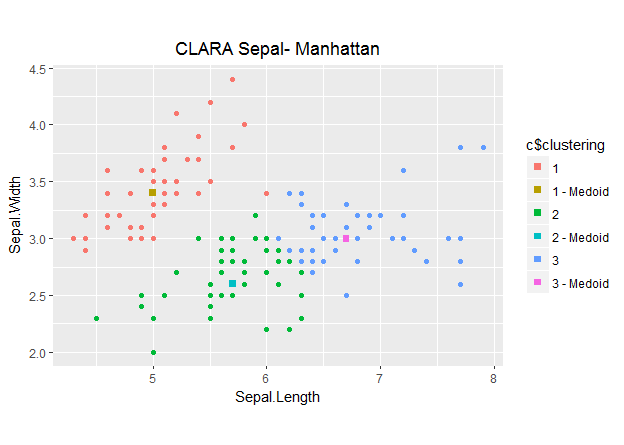
Resultados con CLARA



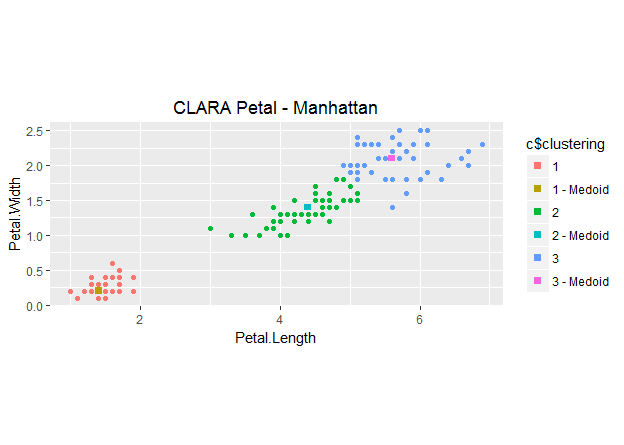
Error: 0.4144793



Error: 0.3685878



Error: 0.538



Error: 0.454

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 (y que, por tanto, son “virtuales”, no existen realmente en el dataset), los centros de PAM son los objetos más centrados en el clúster, lo que es la mediana (y 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.

K-means quiere minimizar la **suma de cuadrados**.

CLARA es una variación de PAM usada para datasets grandes.

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