

# Agrupamiento con Matlab

Partimos de que tenemos los datos en una matriz X Cada fila es un punto de la muestra; cada columna es el valor de una variable.

## Jerárquico

1. Obtenemos sus distancias con `Y=pdist(X);`
2. Hacemos el agrupamiento con `Z=linkage(Y);`
3. Evaluar los agrupamientos: `inconsistent(Z)`  
Hay que fijarse en la cuarta columna. Habrá valores muy bajos y otros altos. Localiza un umbral que separe con la mayor claridad posible
4. Crear el agrupamiento: `T = cluster(Z, 'cutoff', umbral);`  
T es una columna con tantas filas como puntos de la muestra, en la que se indica el número de grupo que le corresponde

## Medias

1. Vamos a iterar variando el número de grupos, para encontrar la cantidad mejor
  1. Crea el agrupamiento con `[idx,C,sumdist] = kmeans(X,numgrupos,'Distance','cityblock);`  
idx es la columna que tiene por cada fila el número de grupos que le corresponde
  2. Evalua el agrupamiento con:  
`[silh,h] = silhouette(X,idx,'cityblock');`  
`xlabel('Factor de distancia')`  
`ylabel('Grupo')`  
En el gráfico que aparece habrá una zona de barras para cada grupo. En un buen agrupamiento deben ser todas, o las más posibles, bastante altas (mayores que 0.5)