

## Lab. 4 - Clustering di dati medici

**Componenti gruppo:**

- Federico Caldart , matricola: 1211144
  - Stefano Panozzo, matricola: 1211143
  - Davide Zago, matricola: 1211260
- 

**Efficienza**

**Domanda 1**

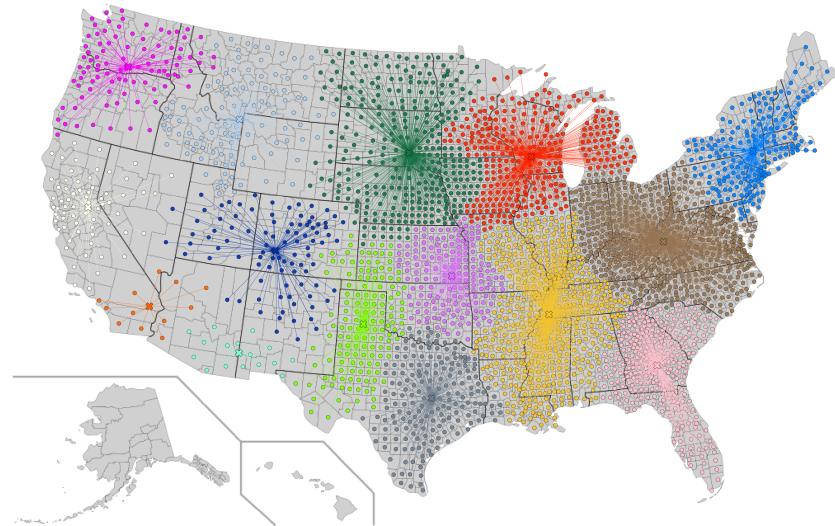


Fig.1: Clustering gerarchico - 3108 contee

## Domanda 2

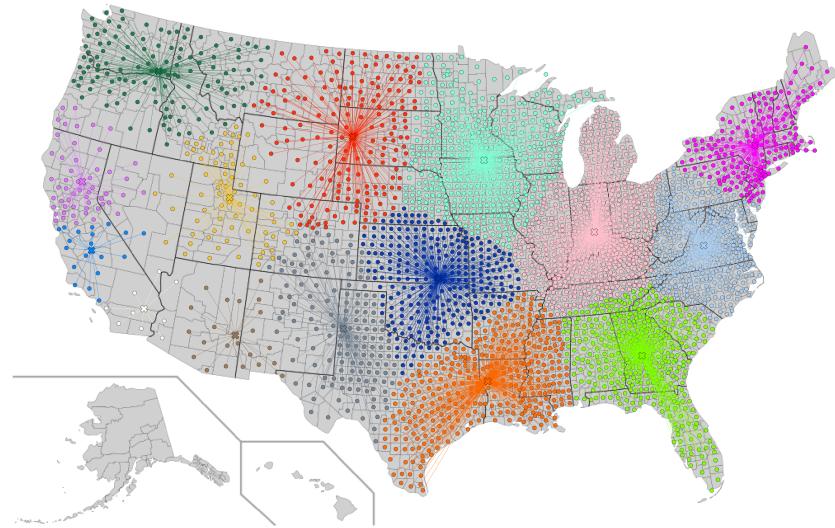


Fig.2: Clustering k-means - 3108 conteee

**Domanda 3**

**Automazione**

**Domanda 4**

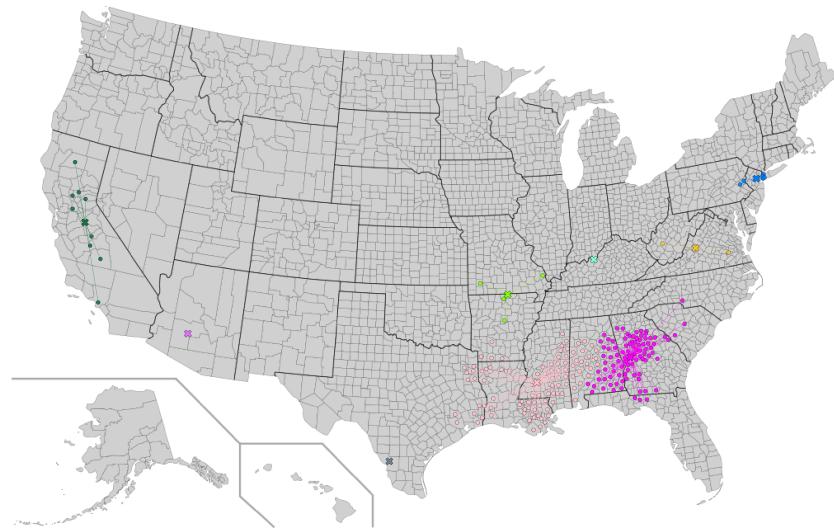


Fig.3: Clustering gerarchico - 212 contee

### **Domanda 5**

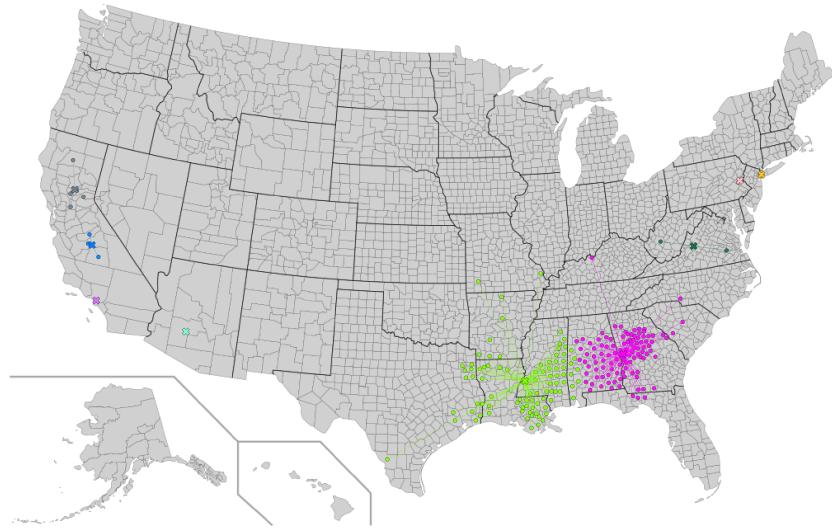


Fig.4: Clustering k-means - 212 contee

### **Domanda 7**

Dalle figure si nota che le contee della costa occidentale con il clustering gerarchico vengono raggruppate in due cluster, di cui uno contiene una sola contea, mentre con il clustering kmeans vengono raggruppati in 4 diversi cluster.

Questo accade perché i centri iniziali di kmeans sono le 15 contee più popolate e 4 di queste sono sulla costa occidentale. Durante le iterazioni del metodo kmeans i quattro centri sulla costa occidentale non riescono a spostarsi e catturano un numero molto basso di punti,

per questo motivo la distorsione di kmeans risulta più elevata.

Per risolvere questo problema si può pensare ad una inizializzazione migliore dei centri che rappresenti meglio la distribuzione del dataset sul piano.

### **Domanda 8**

Il metodo di clustering gerarchico assicura una distorsione relativamente bassa anche senza particolari accorgimenti al contrario di kmeans che richiede una inizializzazione corretta.

## Qualità

### Domanda 9

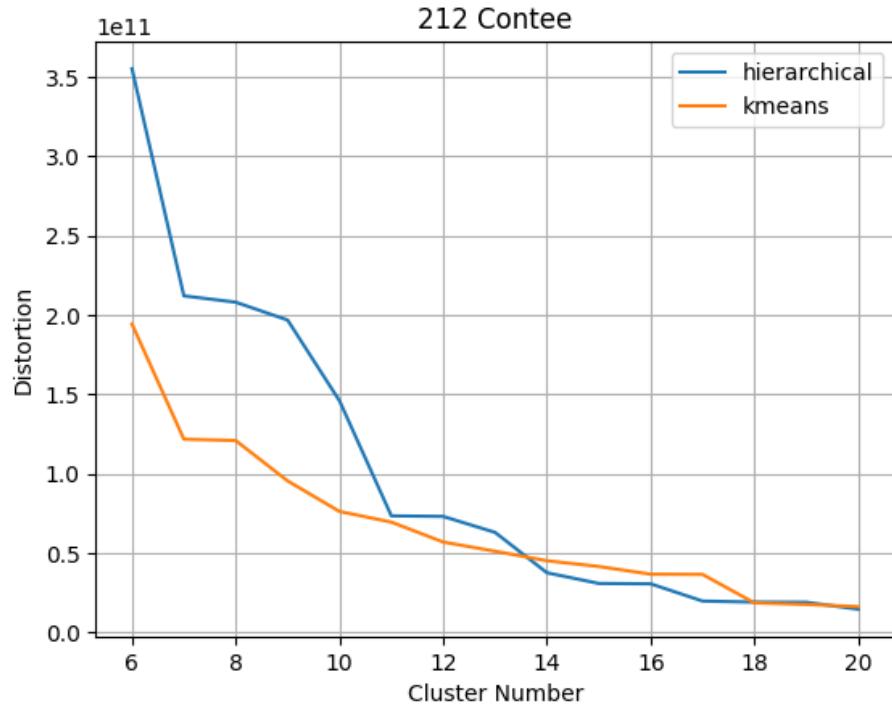


Fig.5: Confronto distorsione - 212 contee

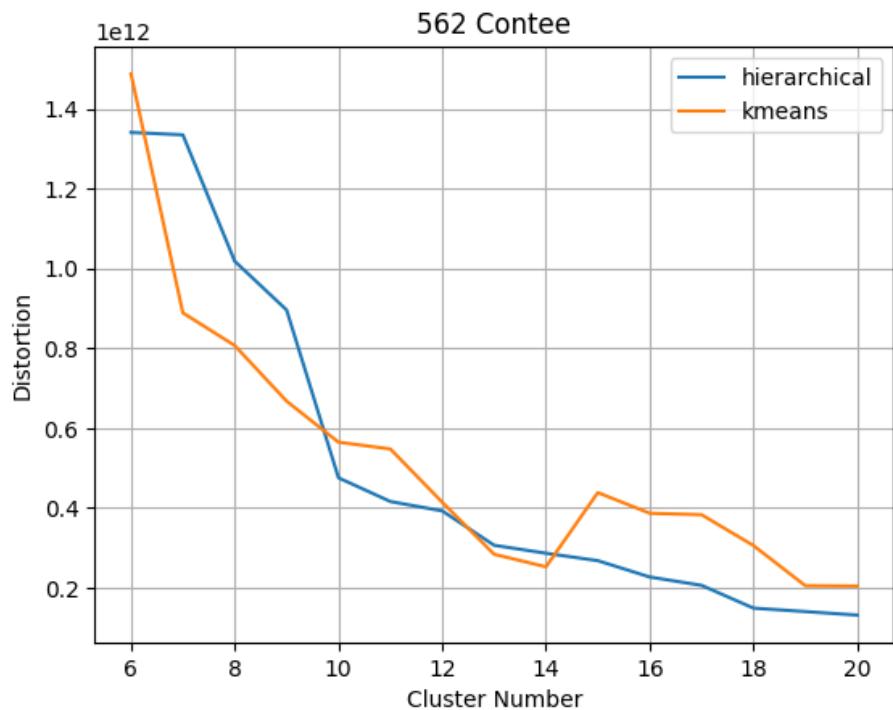


Fig.6: Confronto distorsione - 562 contee

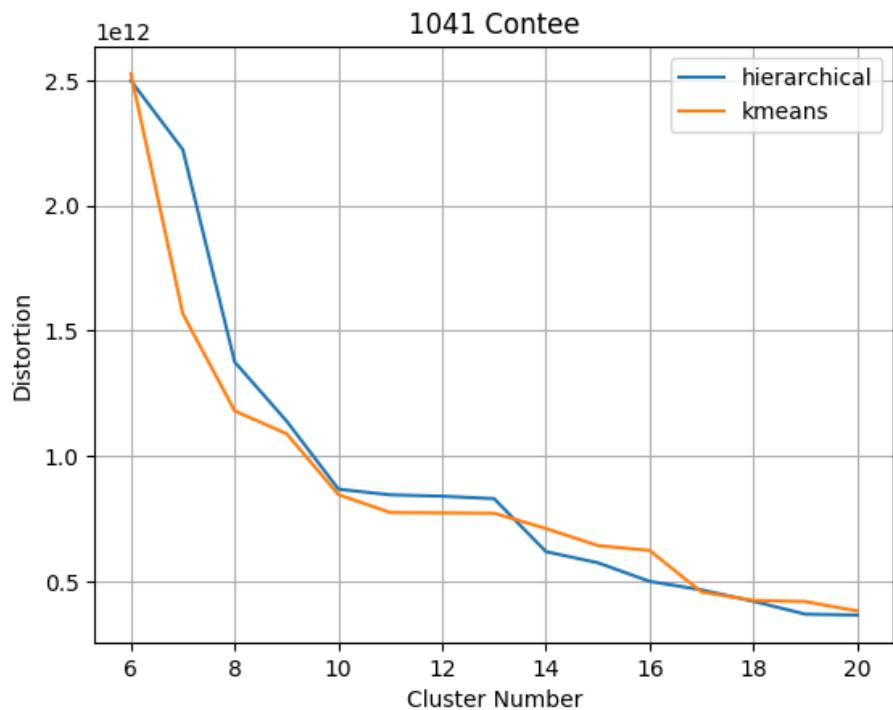


Fig.7: Confronto distorsione - 1041 contee