

# ***DATA SCIENCE***

---

## **AULA 7 - Aprendizagem Não- supervisionada III**

**Prof. Gabriel Resende Machado**



[gabrielmachado@unifeso.edu.com](mailto:gabrielmachado@unifeso.edu.com)



<https://www.linkedin.com/in/machadogabriel>

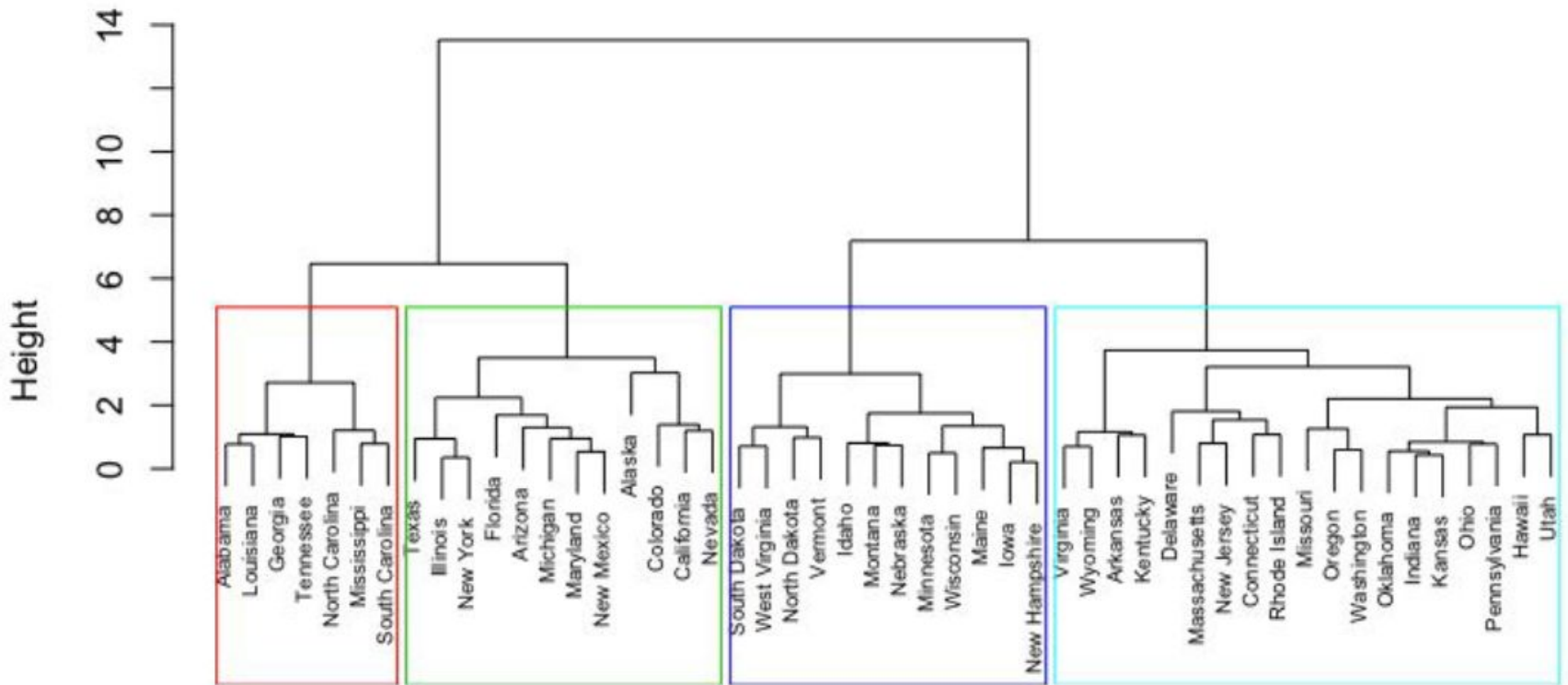


<https://github.com/UNIFESO-Gabriel/data-science>

# Clusterização Hierárquica

- Método de aprendizado não supervisionado por agrupamento;
- Não necessita de uma pré-definição do número de clusters;
- Fornece uma representação visual dos grupos a partir de um dendograma;
- O dendograma representa a relação entre os objetos e mostra quais objetos estão agrupados em diferentes níveis.

# Clusterização Hierárquica - Dendograma



# Clusterização Hierárquica

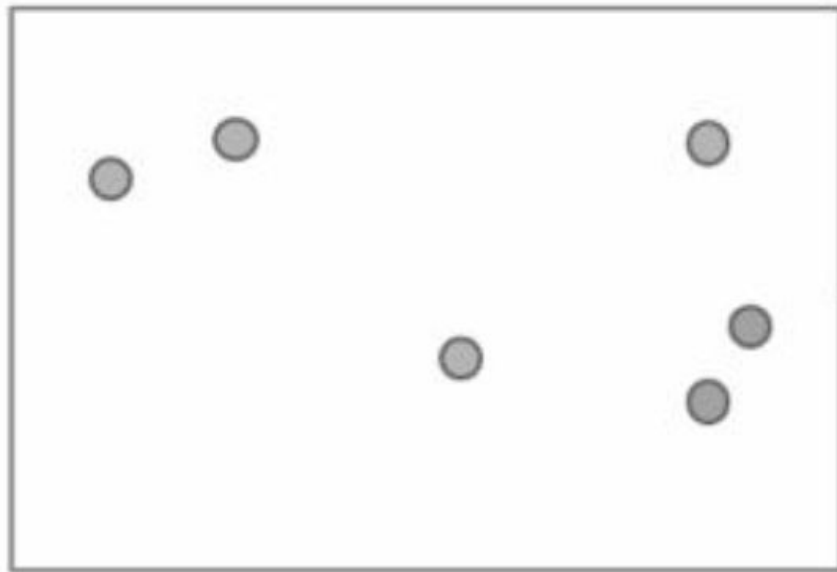
## *Algoritmo*

1. Considere cada elemento do conjunto de treinamento como um *cluster*, ou grupo;
2. Identifique os dois grupos mais próximos entre si;
3. Una dos grupos criando um novo grupo;
  - Repita os passos 2 e 3 até que exista só um único grupo.

# Clusterização Hierárquica

## *Algoritmo*

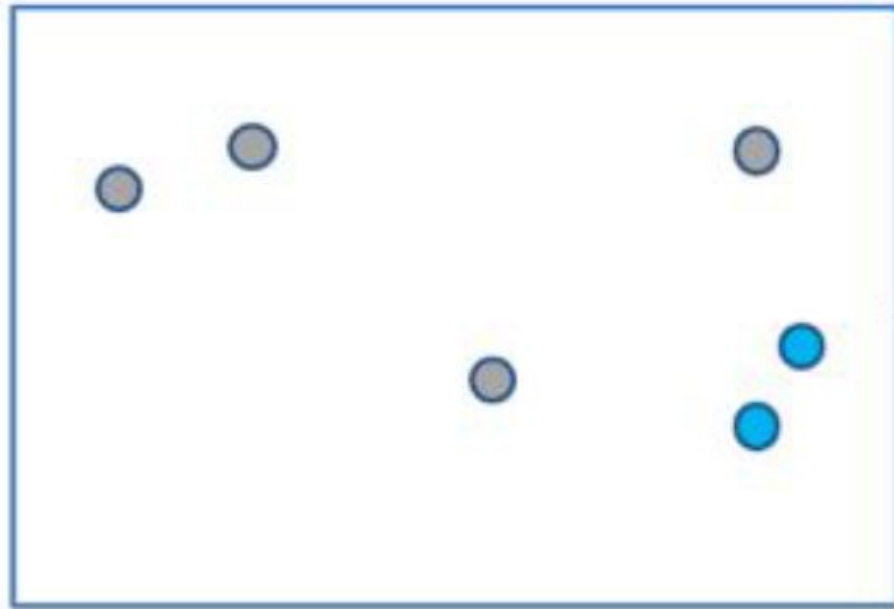
- Identificar e unir os dois grupos mais próximos.



# Clusterização Hierárquica

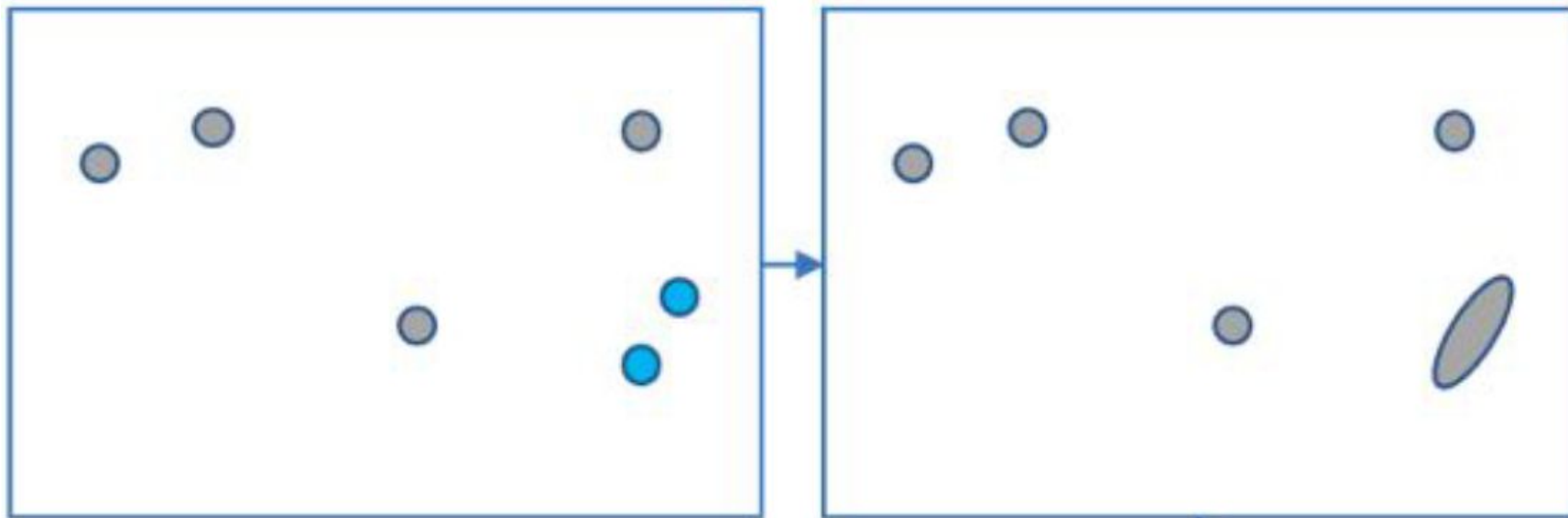
## *Algoritmo*

- Identificar e unir os dois grupos mais próximos.



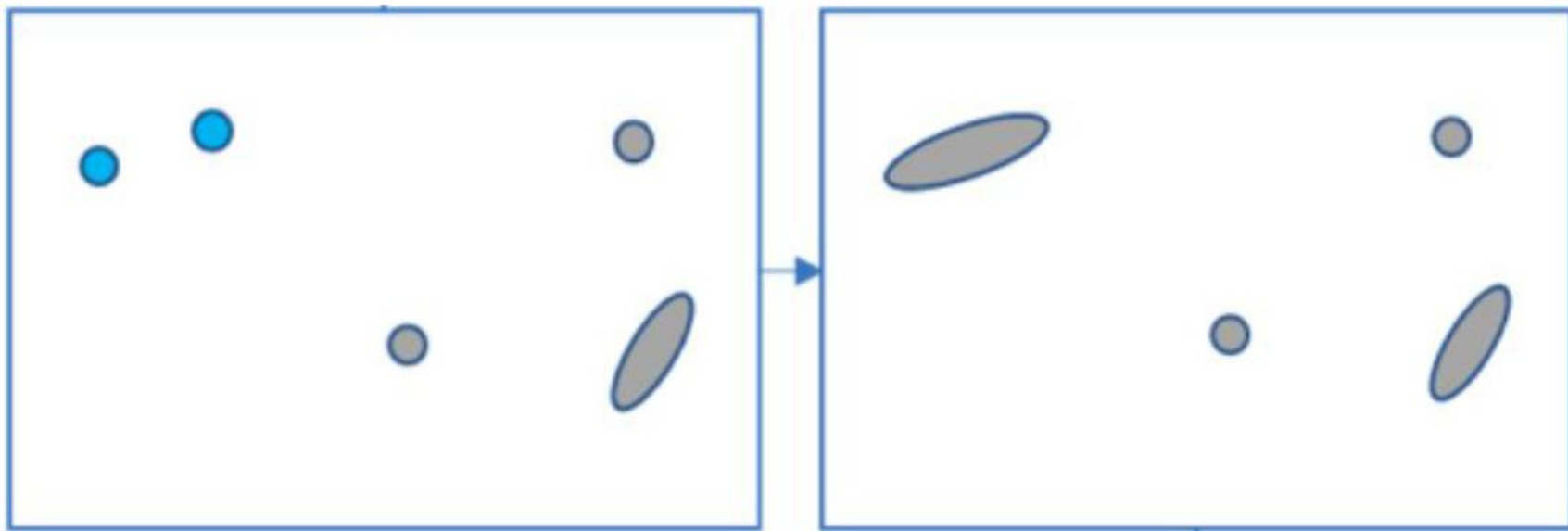
# Clusterização Hierárquica

## *Algoritmo*



# Clusterização Hierárquica

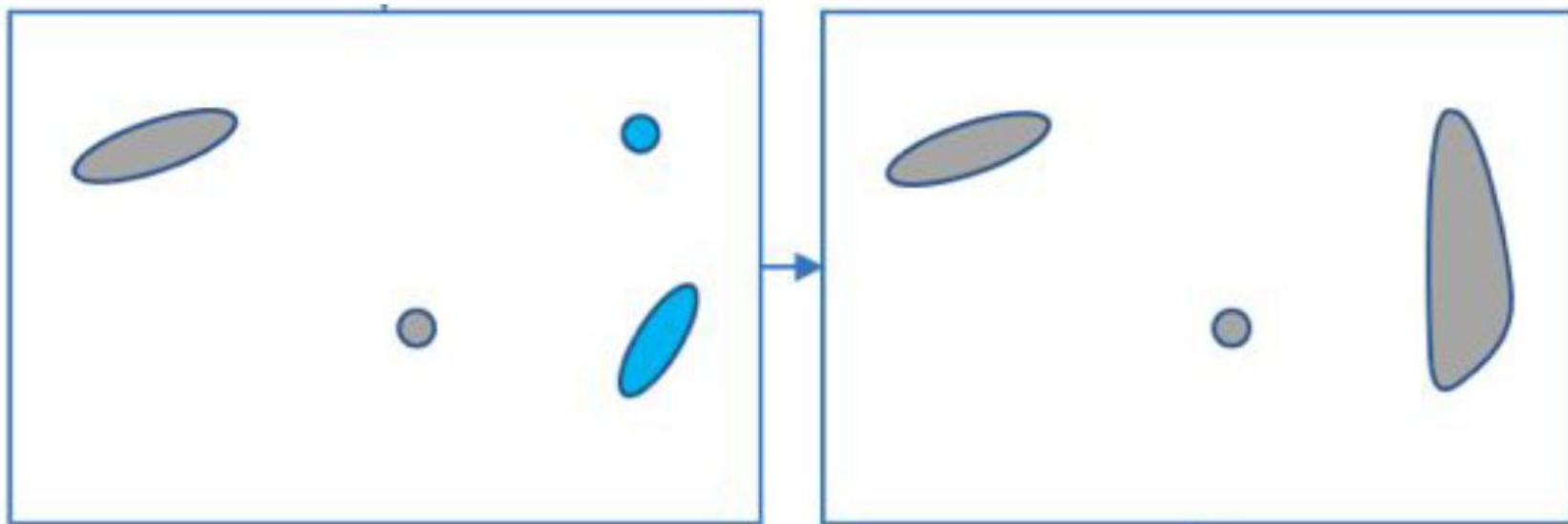
## *Algoritmo*





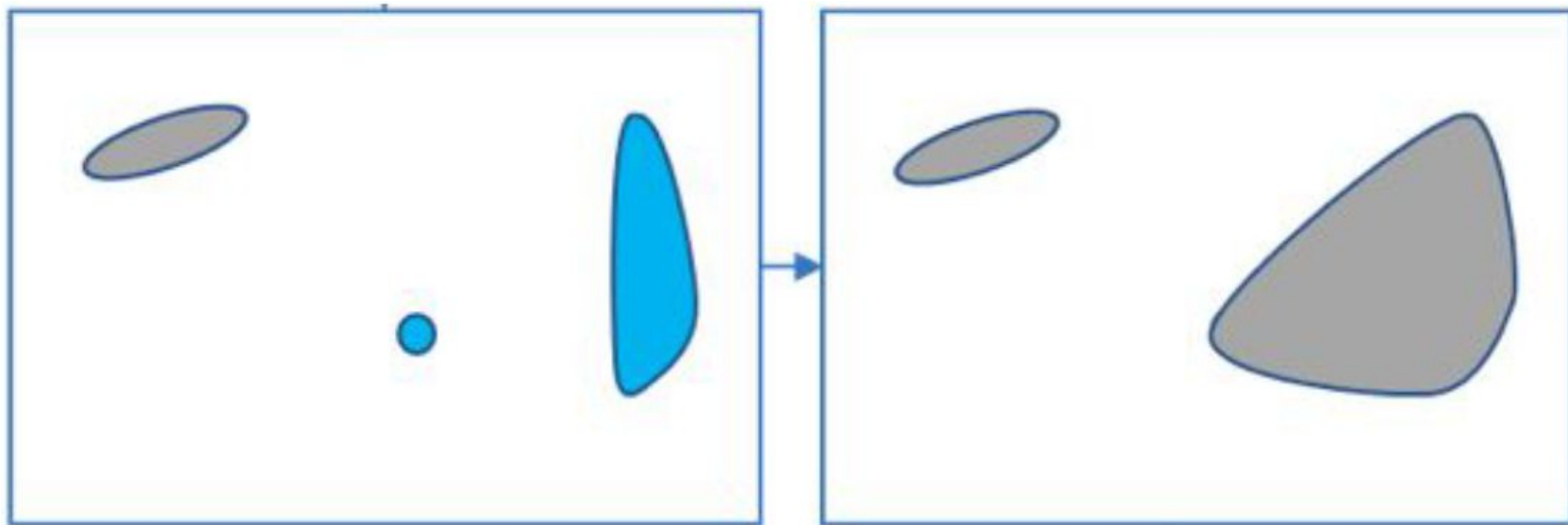
# Clusterização Hierárquica

## *Algoritmo*



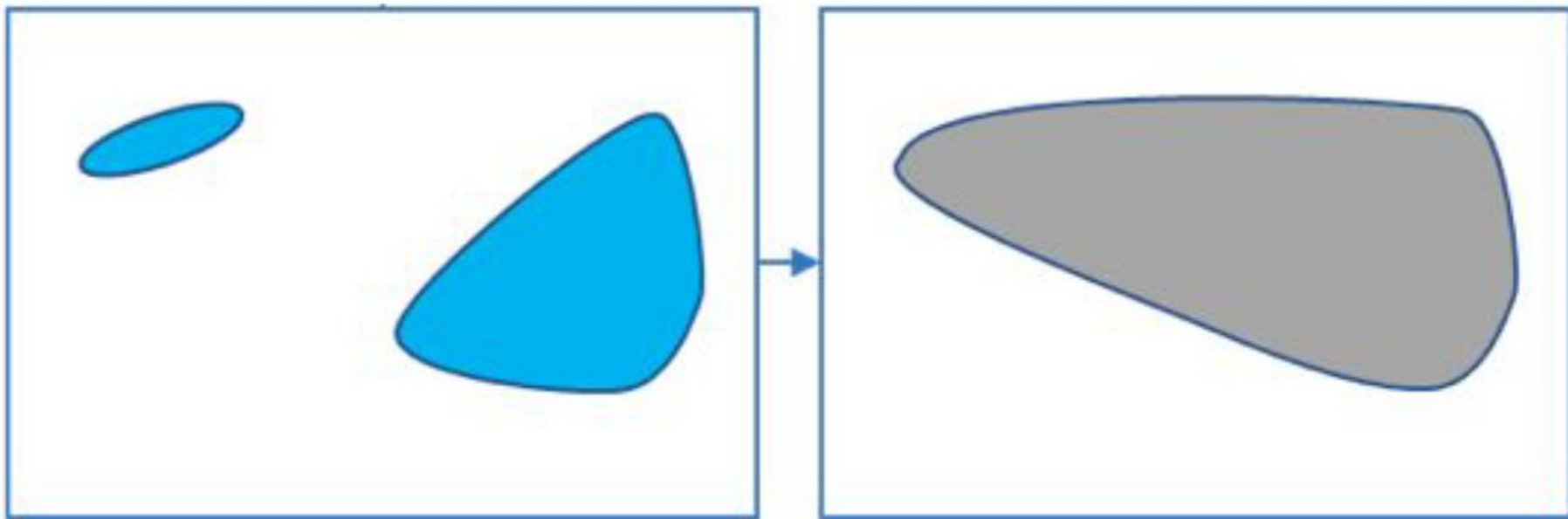
# Clusterização Hierárquica

## *Algoritmo*



# Clusterização Hierárquica

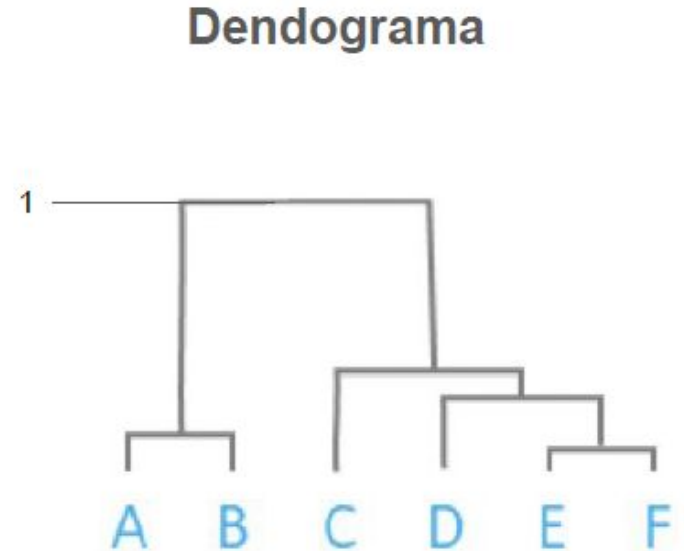
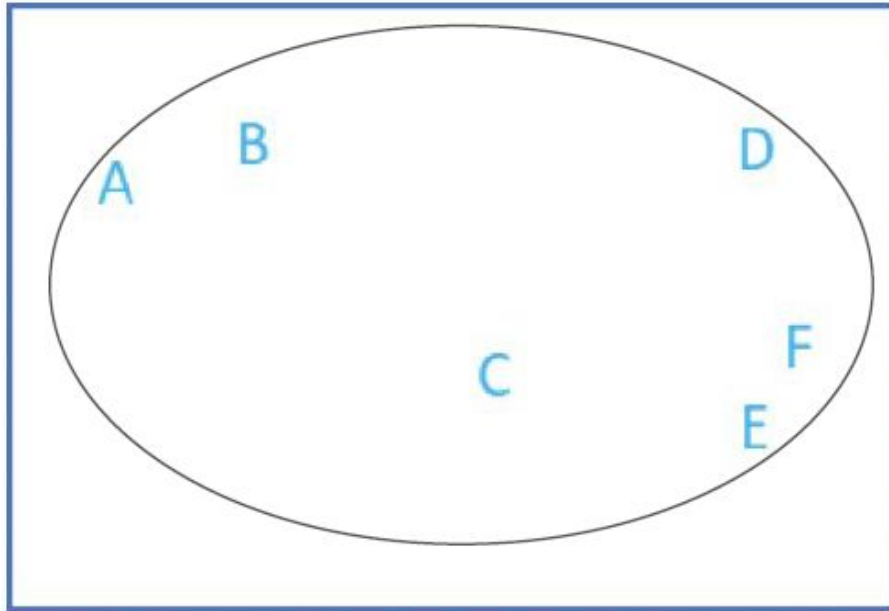
## *Algoritmo*



# Clusterização Hierárquica

## *Algoritmo*

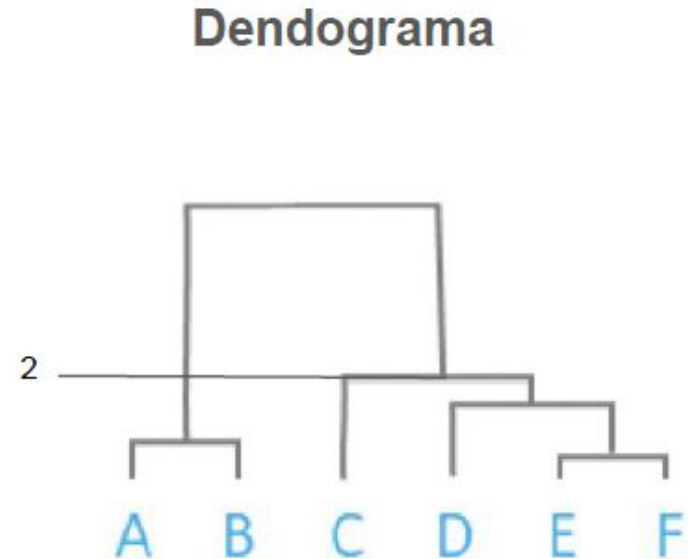
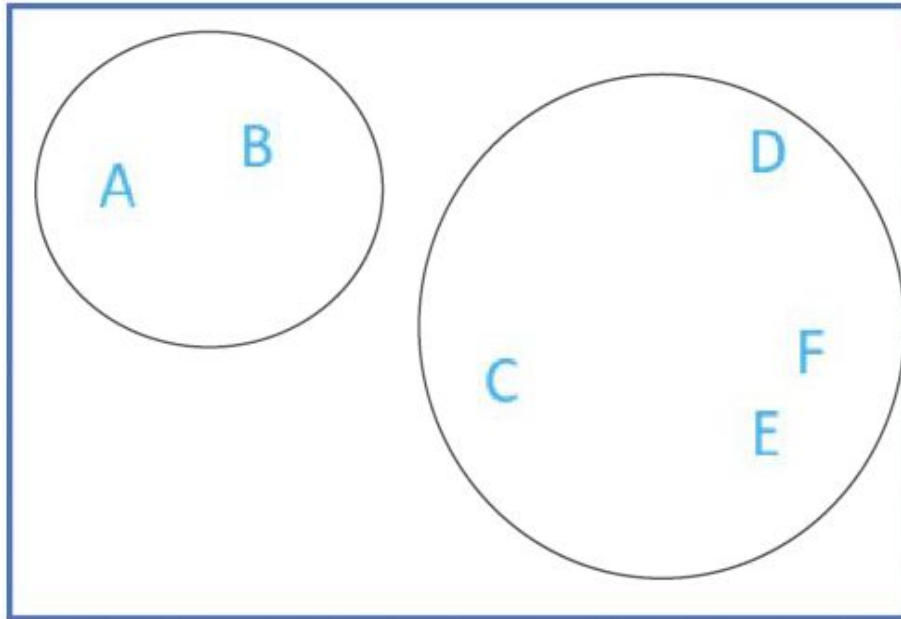
- Resultado: dendograma representando os grupos.



# Clusterização Hierárquica

## *Algoritmo*

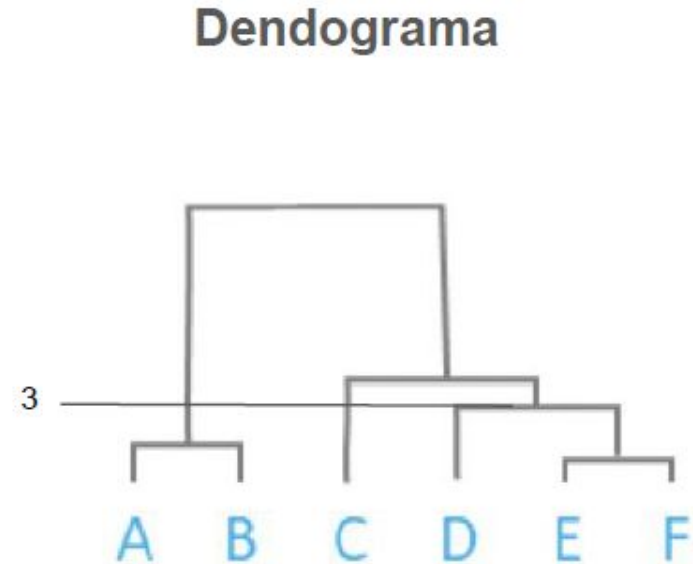
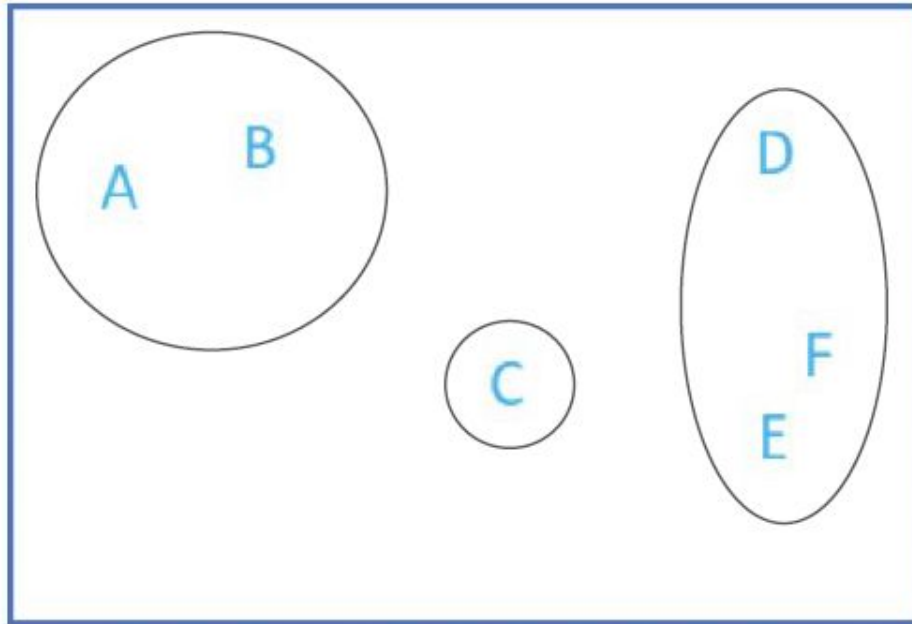
- Resultado: dendograma representando os grupos.



# Clusterização Hierárquica

## *Algoritmo*

- Resultado: dendograma representando os grupos.



# Clusterização Hierárquica

- **Clusterização Aglomerativa (*bottom-up*)**: inicia com  $n$  grupos, sendo  $n$  o número de itens a serem agrupados. O procedimento termina quando apenas um grupo permanecer.
  - Agrupa grupos similares.
- **Clusterização Divisiva (*top-down*)**: inicia com um único grupo. O procedimento termina quando são identificados  $n$  grupos, sendo  $n$  o número total de itens a serem agrupados;
  - Divide grupos heterogêneos.

# Clusterização Hierárquica

- Resultado reproduzível;
- Alto custo computacional;
- Escalonamento e tratamento prévio de outliers é recomendado.



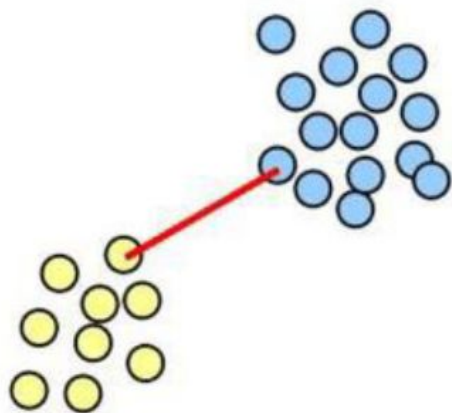
# Clusterização Hierárquica

## *Métricas de distância*

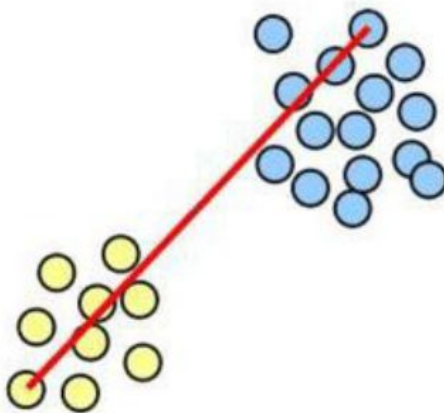
- A forma como é medida a dissimilaridade (distância) dos grupos influencia o resultado final do algoritmo.
  - **Clusterização por Ligação Mínima:** distância mínima entre dois pontos quaisquer dos clusters (*single linkage*);
  - **Clusterização por Ligação Máxima:** distância máxima entre dois pontos quaisquer dos clusters (*complete linkage*);
  - **Clusterização por Ligação Média:** distância média entre dois pontos quaisquer dos clusters (*average linkage*).

# Clusterização Hierárquica

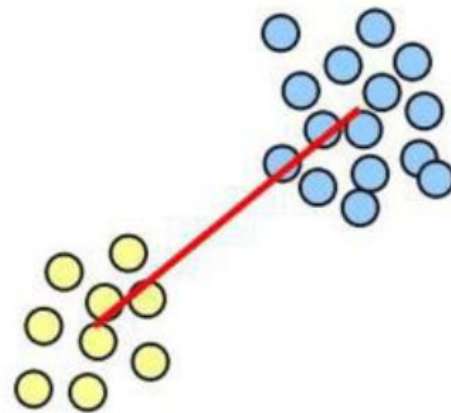
## *Métricas de distância*



single-link



complete-link

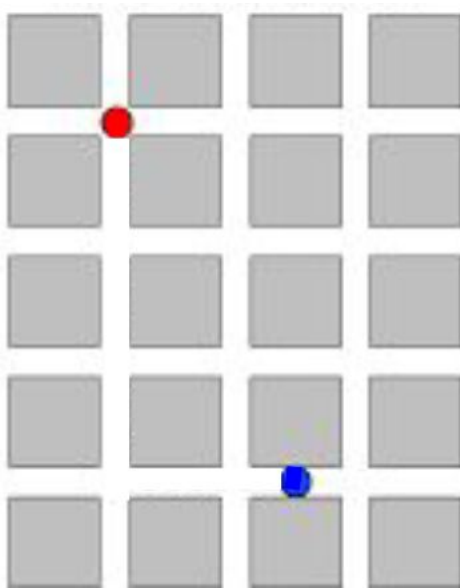


average-link

# Clusterização Hierárquica

## *Algoritmo*

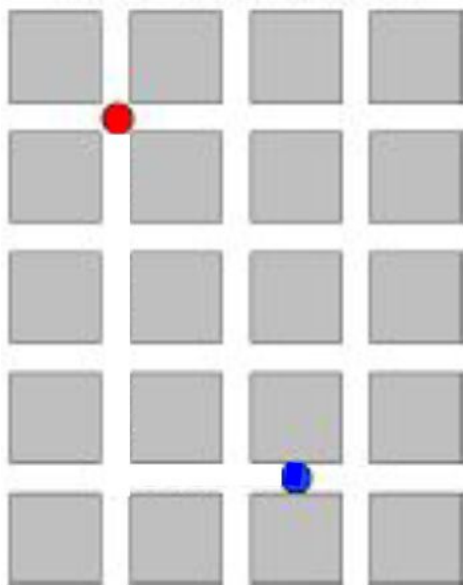
- Como medir a distância entre dois pontos?.



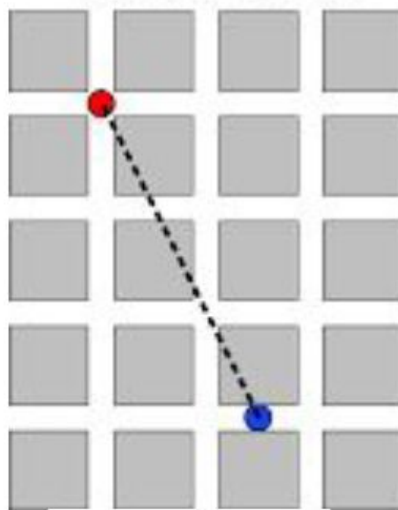
# Clusterização Hierárquica

## *Algoritmo*

- Como medir a distância entre dois pontos? .

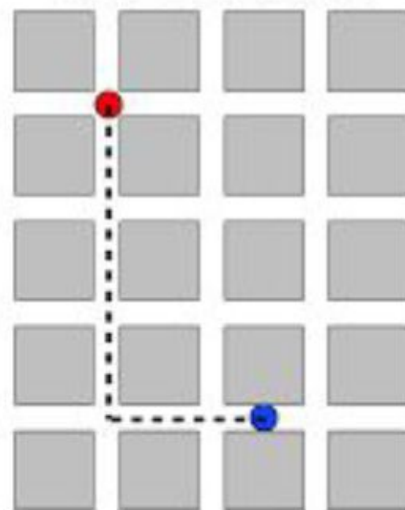


Distância Euclidiana



$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

Distância de Manhattan



$$d(x, y) = \sum_{i=1}^n |y_i - x_i|$$

# Clusterização Hierárquica

## *Exercício 1*

Faça a clusterização hierárquica dos estados americanos a partir dos dados de altura e peso de seus habitantes.

Link do Colab: <https://shorturl.at/ksIKM>.



# Clusterização Hierárquica

## *Exercício 2*

Faça a clusterização hierárquica dos pokémons a partir dos dados de peso e velocidade. Ao final, exiba o dendograma para os 10 primeiros pokémons.

Link do Colab: <https://shorturl.at/KPZ16>.



# ***DATA SCIENCE***

---

## **AULA 7 - Aprendizagem Não-Supervisionada III**

**Dúvidas e/ou perguntas?**



[gabrielmachado@unifeso.edu.com](mailto:gabrielmachado@unifeso.edu.com)



<https://www.linkedin.com/in/machadogabriel>



<https://github.com/UNIFESO-Gabriel/data-science>