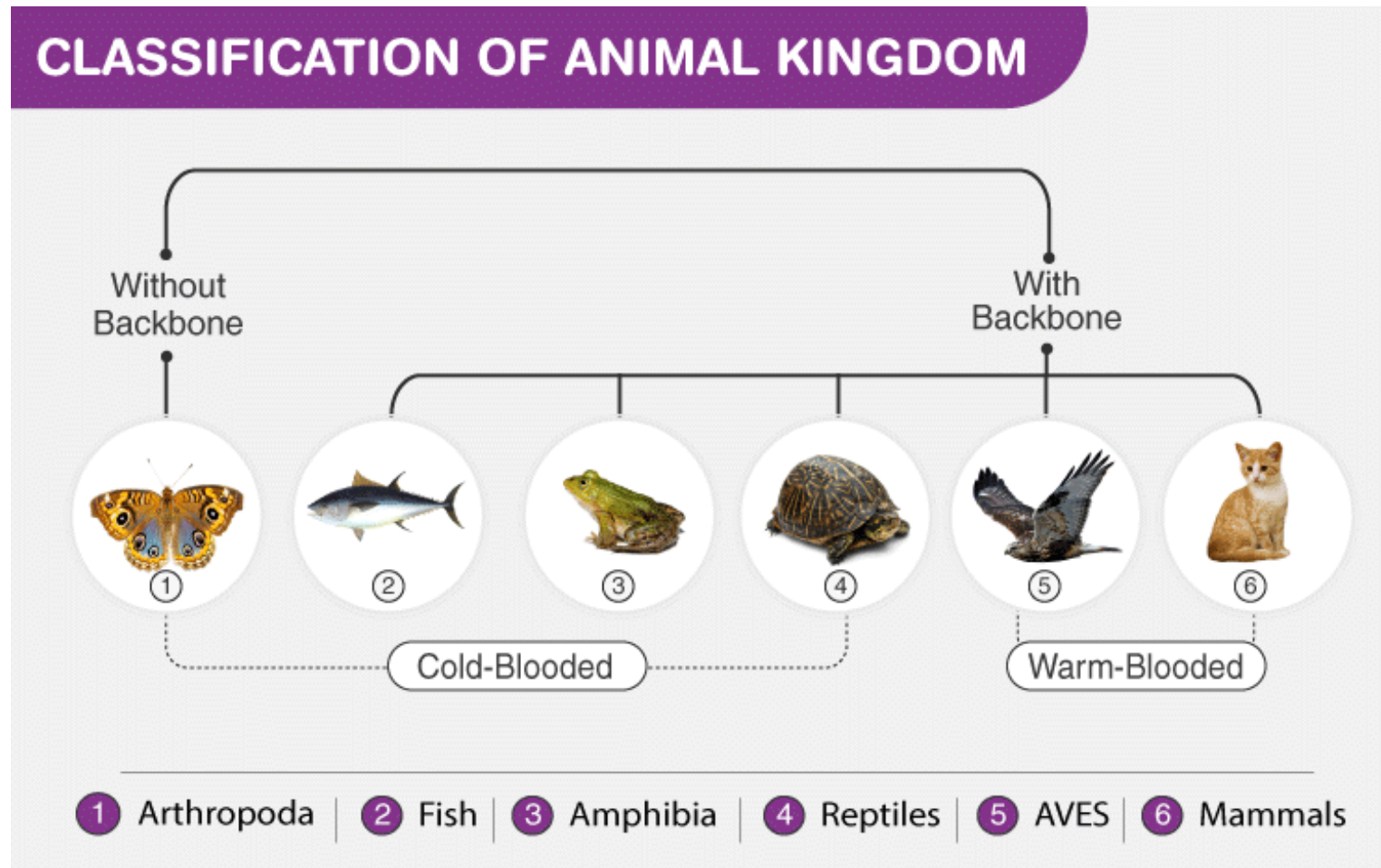


## Introducción a Machine learning aprendizaje no supervisado

Curso de análisis de datos con R  
Asociación Argentina de Bioinformática y Biología Computacional



Todos los animales son iguales, pero algunos animales son más iguales que otros

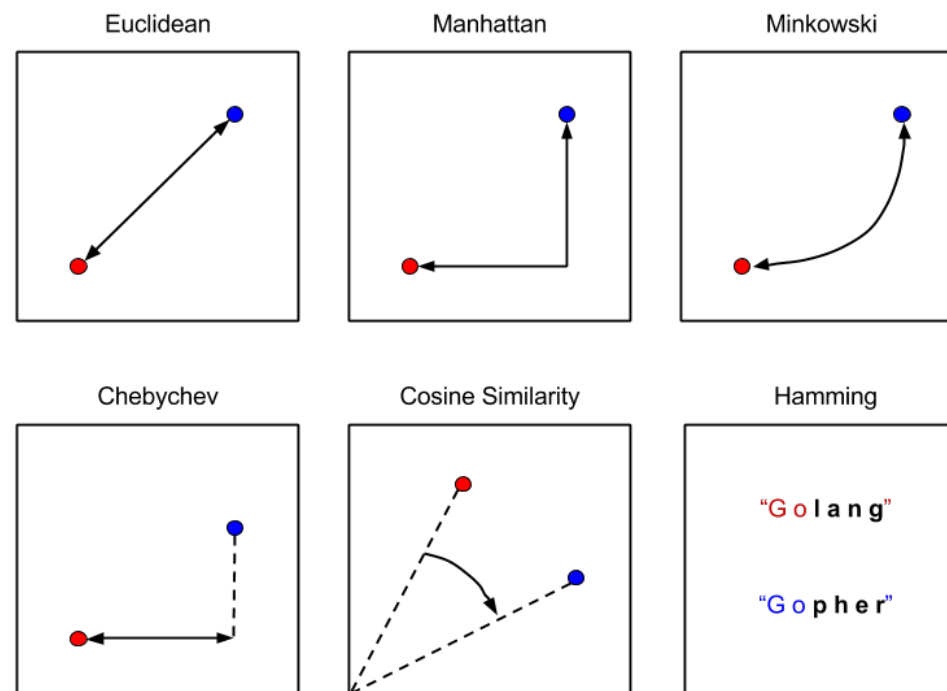


<https://byjus.com/biology/classification-of-animal-kingdom/>

# Distancias

Lo primero que podemos hacer para ver si dos observaciones son similares o no, es calcular una distancia entre ellas.

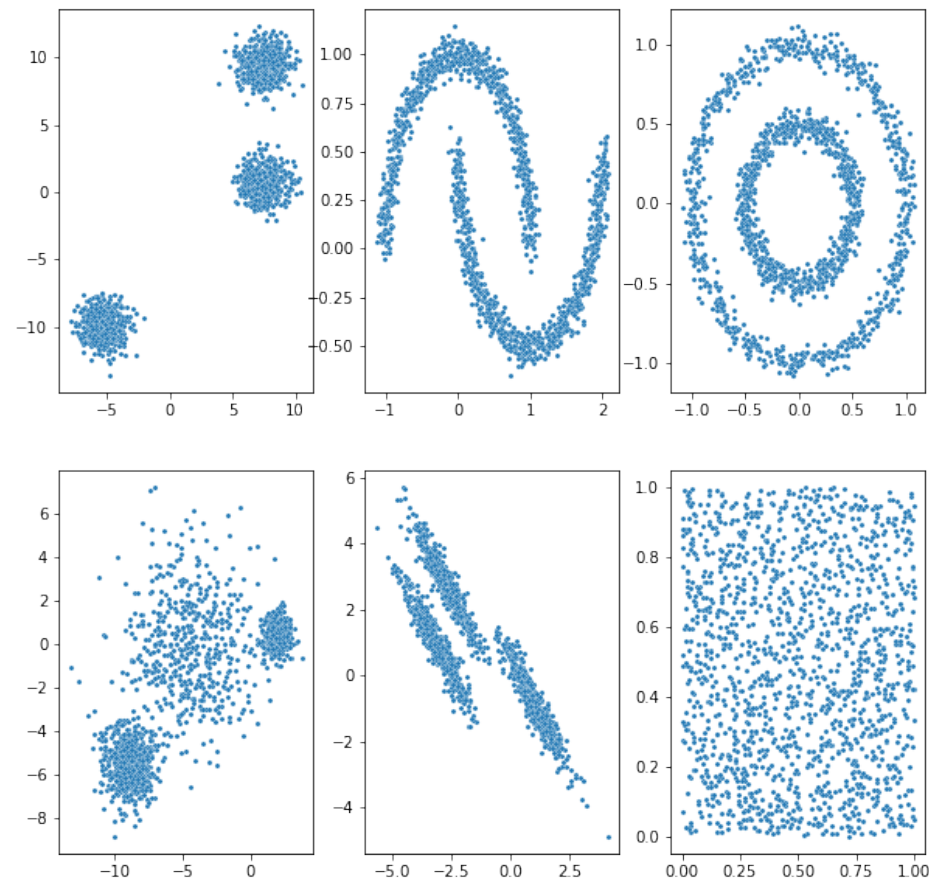
¿Pero todos los espacios son iguales? ¿Qué forma tendrá el espacio en el que viven nuestros datos?



[https://subscription.packtpub.com/book/big\\_data\\_and\\_business\\_intelligence/9781785882104/6/ch06lv11sec40/measuring-distance-or-similarity](https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781785882104/6/ch06lv11sec40/measuring-distance-or-similarity)

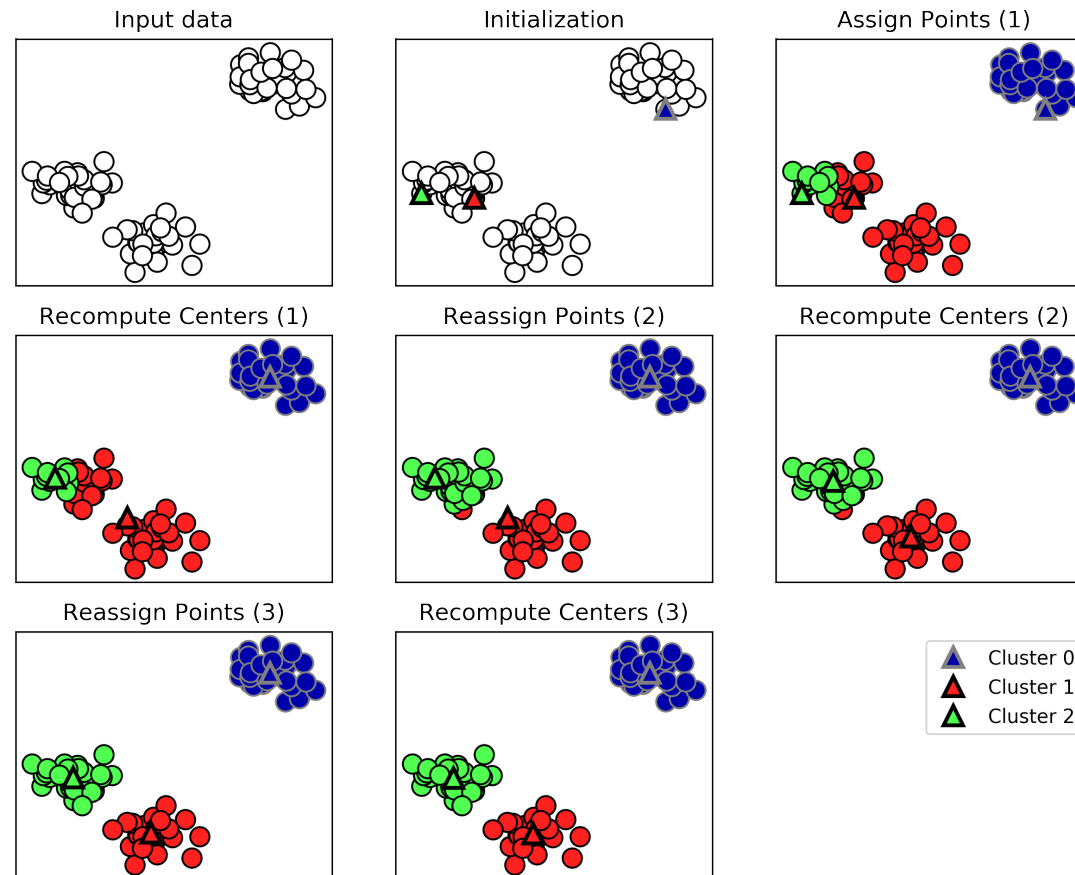
## Ejemplos de juguete

Construyamos algunos ejemplos de juguete para entender mejor el problema.



## K-means

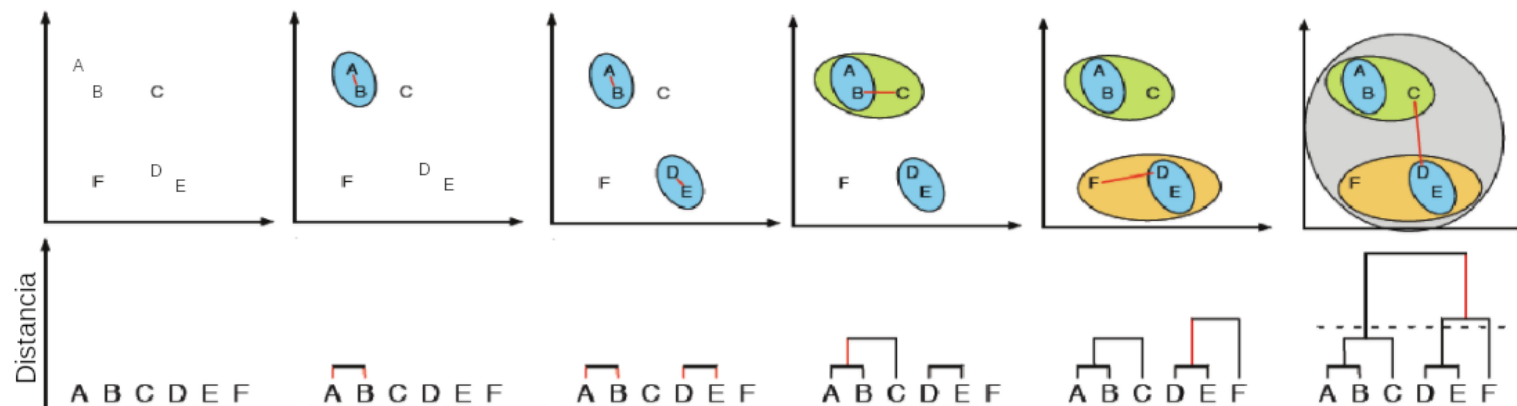
El método más básico de clustering se conoce como *k-means*.



¿Que necesitamos definir para usar esta técnica? ¿Qué problemas puede tener? ¿Qué tipo de clusters podemos detectar?

## Clustering jerárquico

En lugar de dividir todas las observaciones en grupos, agregamos las observaciones en grupos cada vez más grandes hasta juntarlas todas en un único grupo.



¿Que necesitamos definir para usar esta técnica? ¿Qué tipo de clusters podemos detectar?

