

Agrupar y reinarás

daniela rodriguez- instituto de cálculo



Big data - Walter sosa escudero

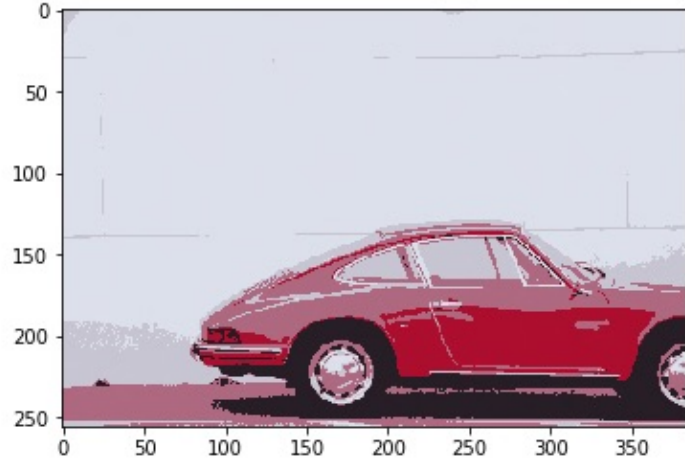
«Cajón de sastre»
es la expresión adecuada
para referirse a un conjunto
de cosas variadas y desor-
denadas'...

Nose dice
«cajón desas-
tre».

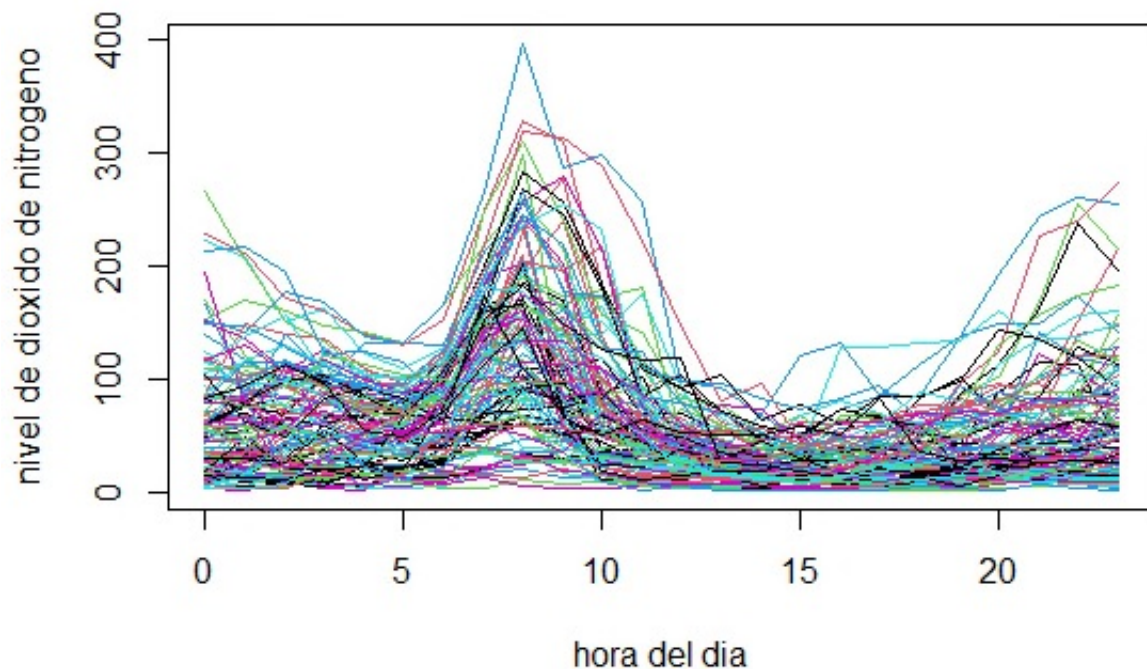
©DonFardino



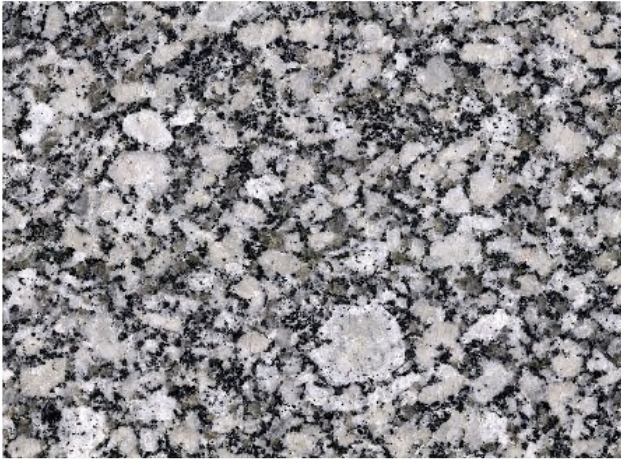
Problemas distintos pero parecidos



Problemas distintos pero parecidos



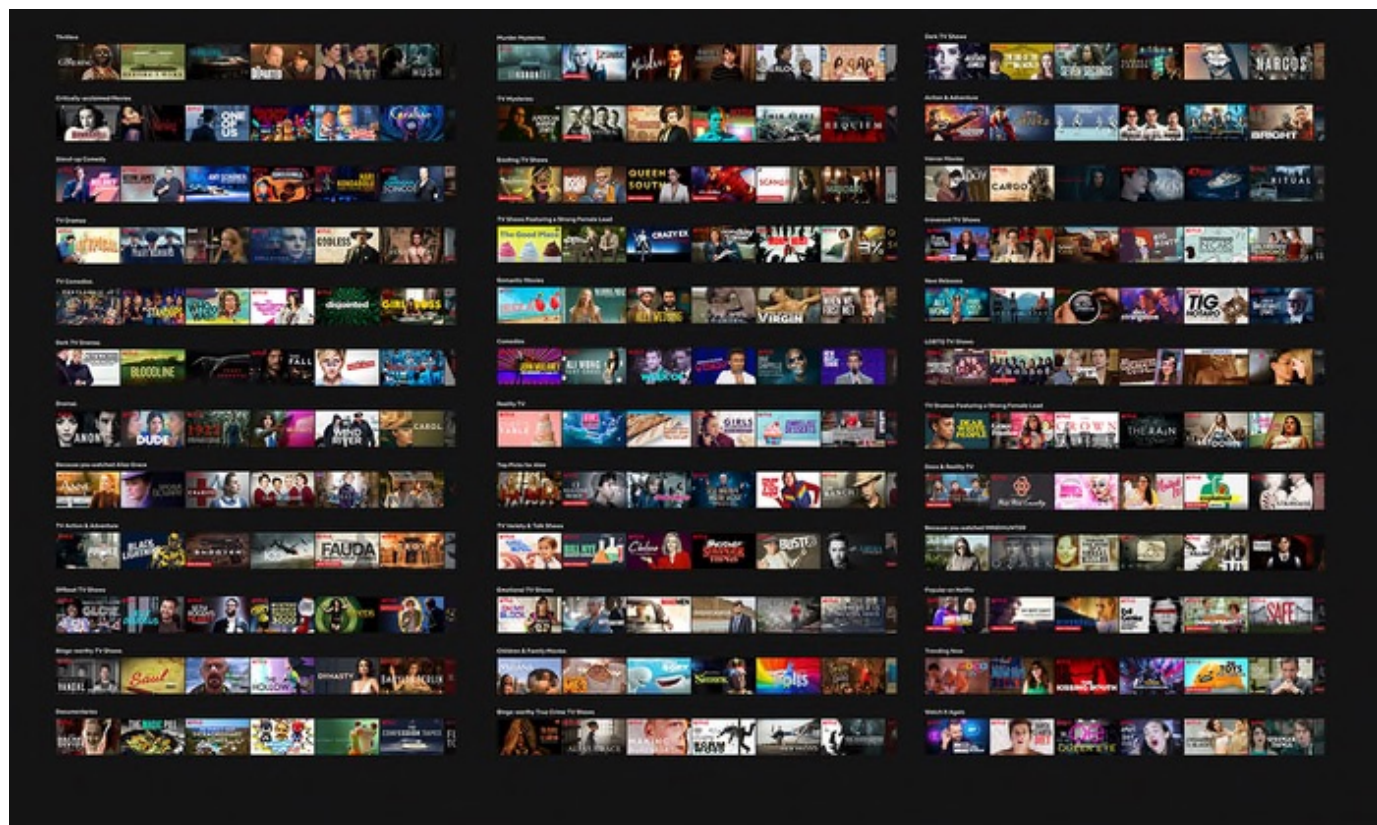
Problemas distintos pero parecidos



Problemas distintos pero parecidos

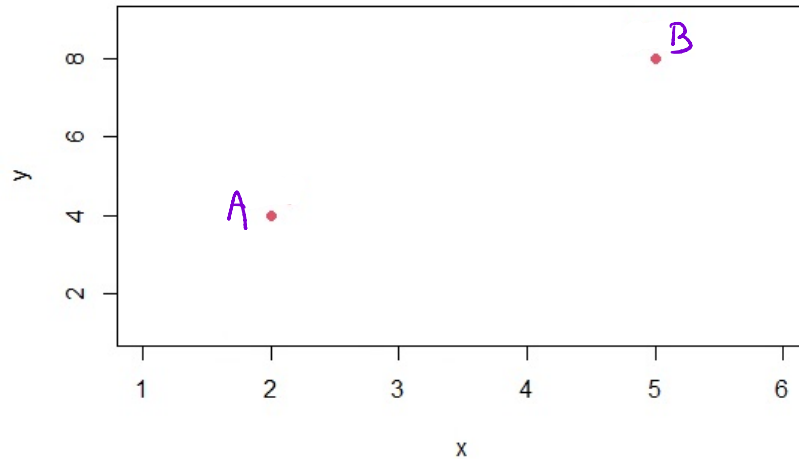


Problemas distintos pero parecidos



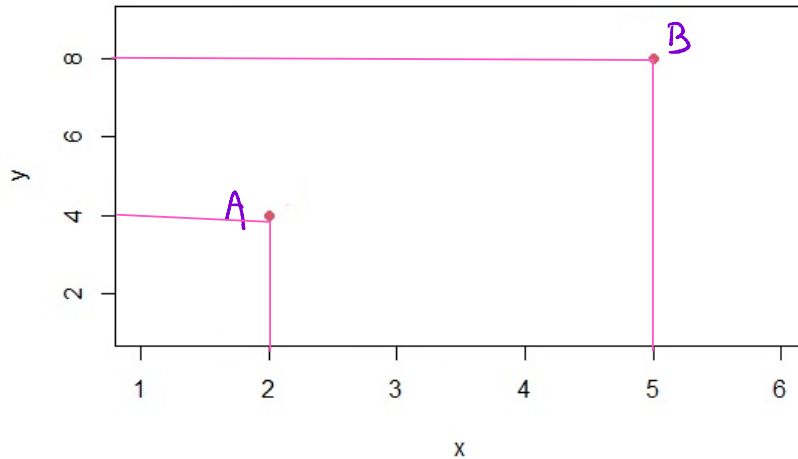
Un poquito de matemática

Cuál es la distancia entre A y B?



Un poquito de matemática

Cuál es la distancia entre A y B?

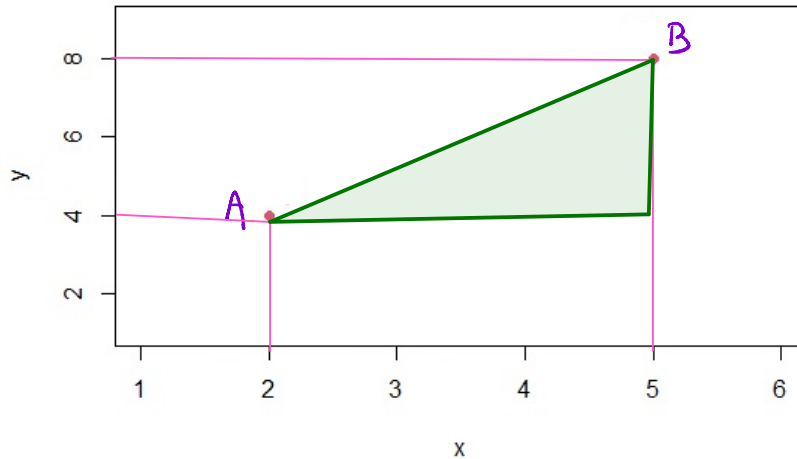


$$A = (2, 4)$$

$$B = (5, 8)$$

Un poquito de matemática

Cuál es la distancia entre A y B?

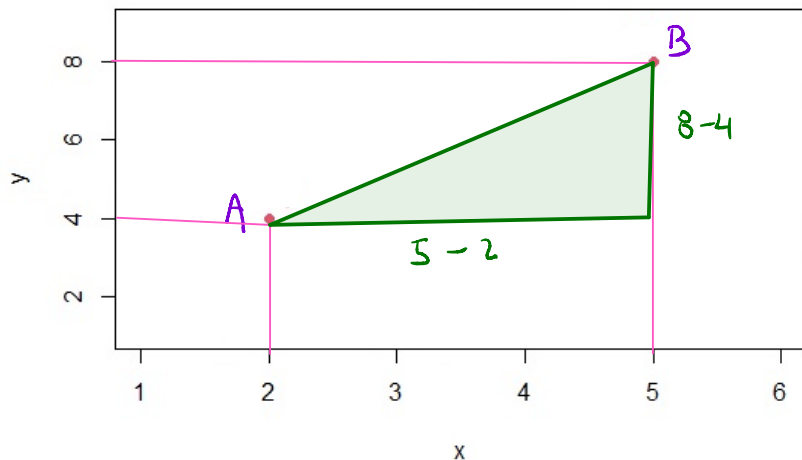


$$A = (2, 4)$$

$$B = (5, 8)$$

Un poquito de matemática

Cuál es la distancia entre A y B?



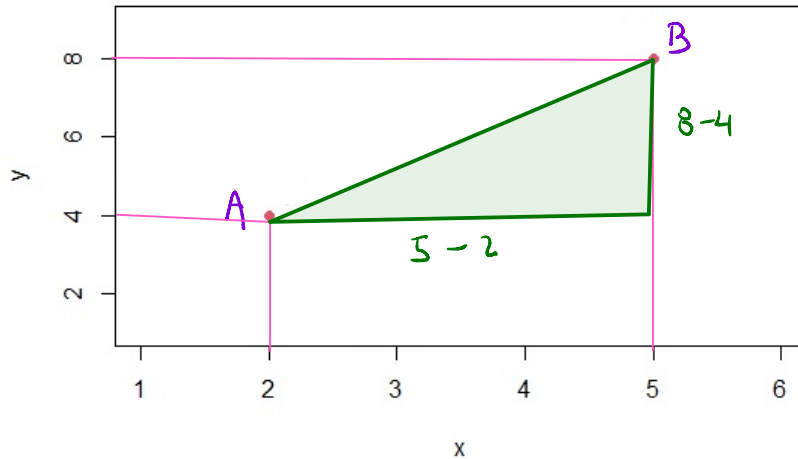
$$A = (2, 4)$$

$$B = (5, 8)$$

$$d(A, B) = \sqrt{(5-2)^2 + (8-4)^2} = \sqrt{9+16} = 5$$

Un poquito de matemática

Cuál es la distancia entre A y B?



$$A = (a_1, a_2)$$

$$B = (b_1, b_2)$$

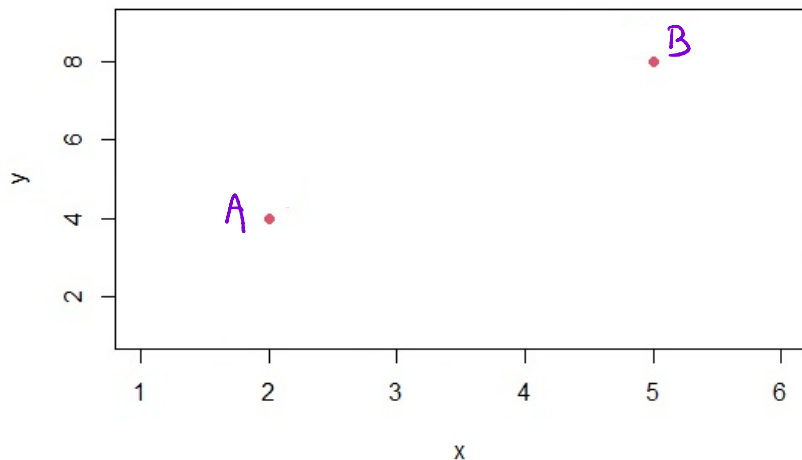
$$A = (2, 4)$$

$$B = (5, 8)$$

$$d(A, B) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}$$

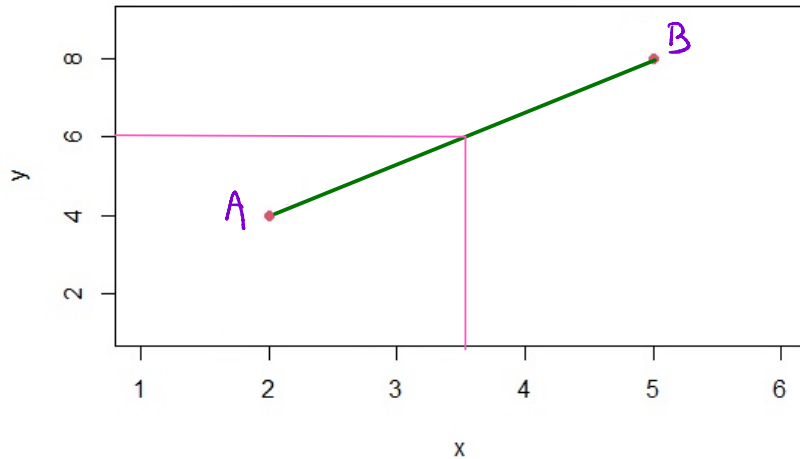
Un poquito de matemática

Cuál es el punto medio entre A y B?



Un poquito de matemática

Cuál es el punto medio entre A y B?



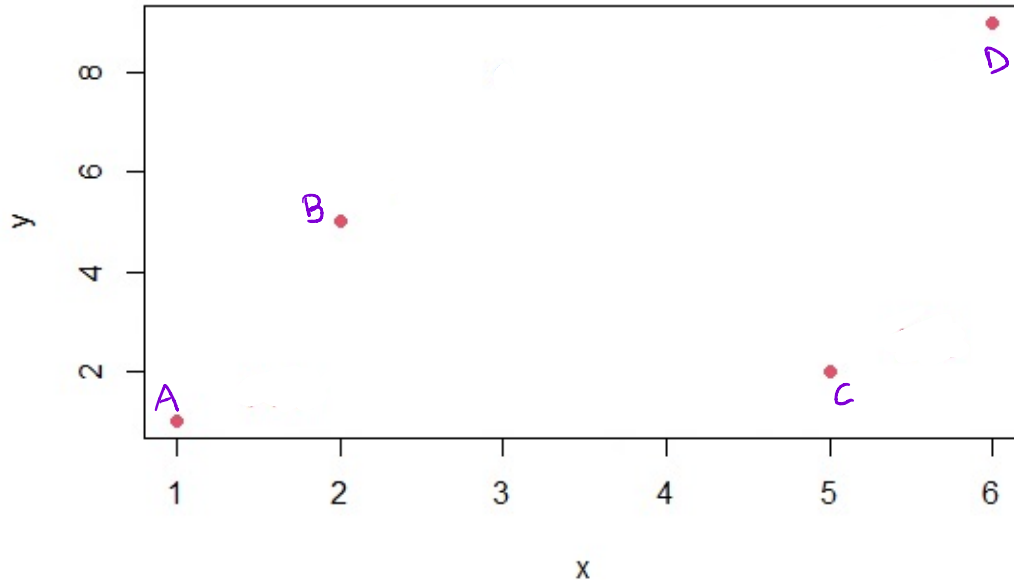
$$A = (a_1, a_2)$$

$$B = (b_1, b_2)$$

$$\overline{AB} = \left(\frac{a_1 + b_1}{2}, \frac{a_2 + b_2}{2} \right)$$

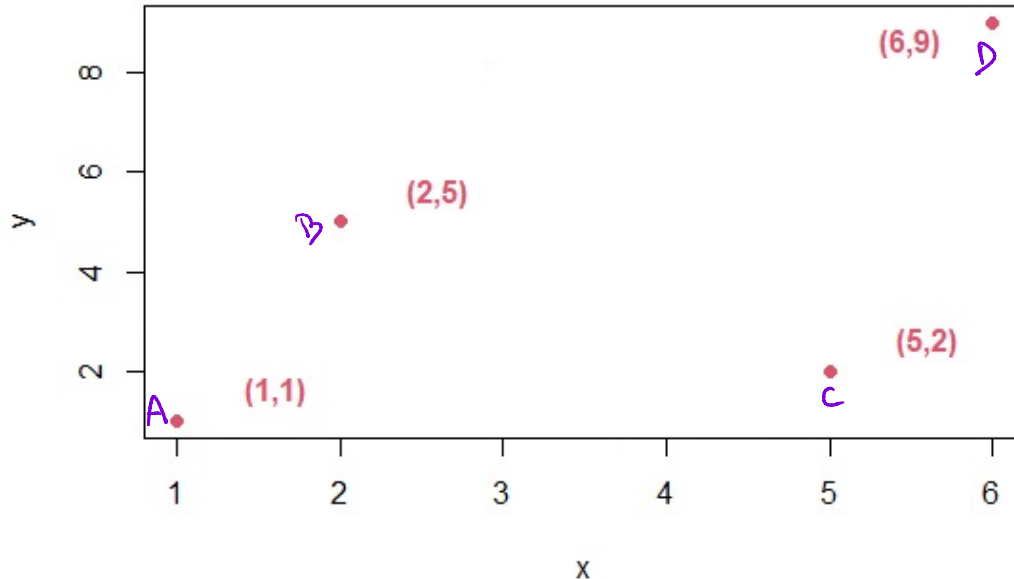
Un poquito más de matemática

Cuál es el punto medio entre A, B, C y D?



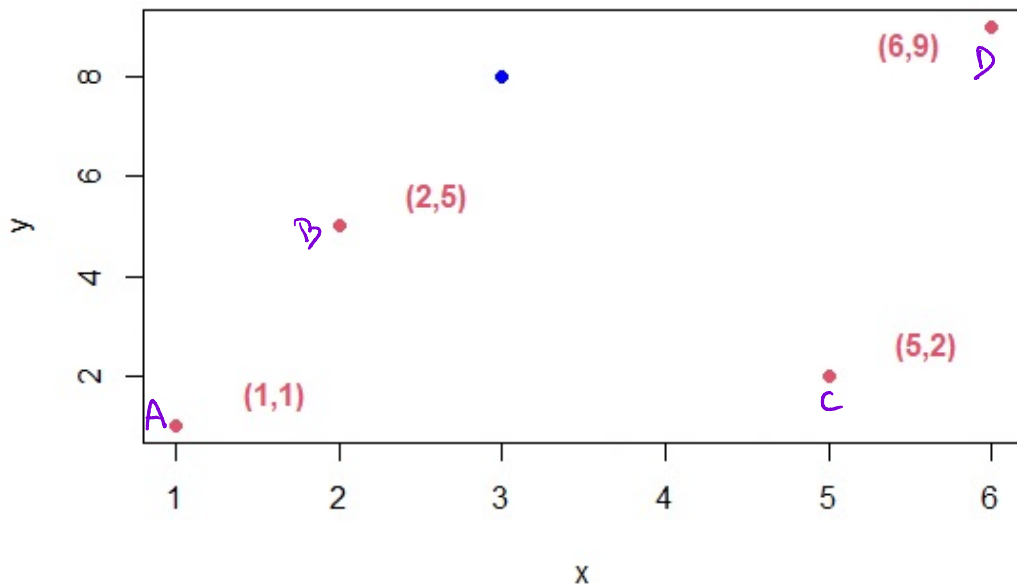
Un poquito más de matemática

Cuál es el punto medio entre A, B, C y D?



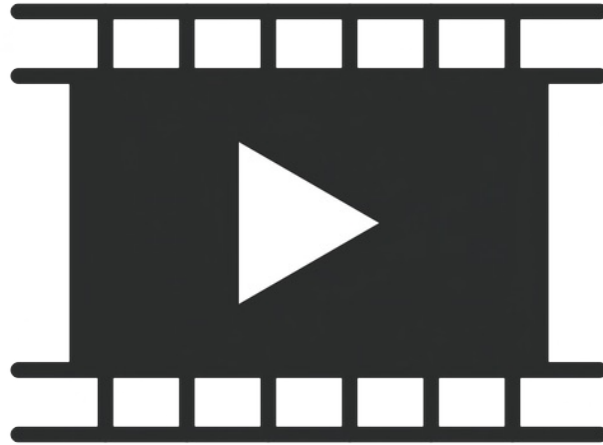
Un poquito más de matemática

Cuál es el punto medio entre A, B, C y D?



Un poquito más de matemática

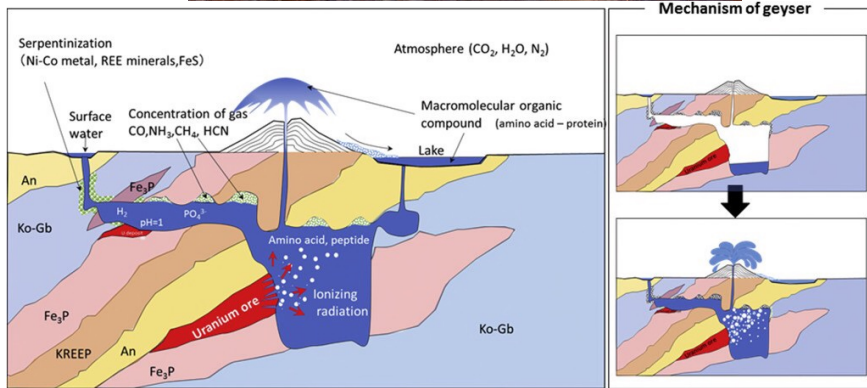
Cuál es el punto medio entre A, B, C y D?



Los géiseres

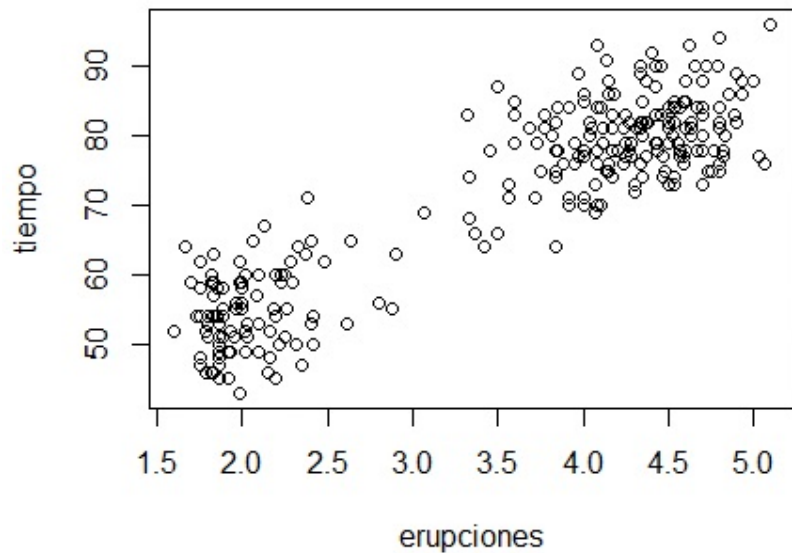


Old Faithful (Yellowstone)



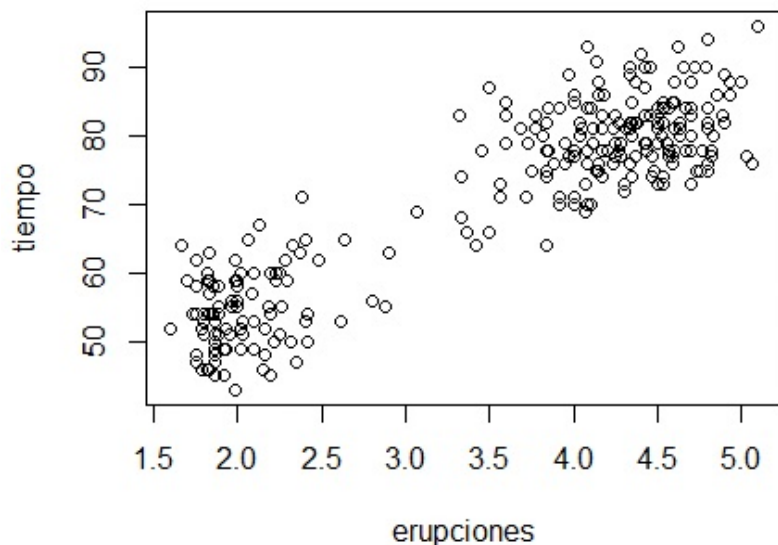
Old Faithful

Datos de Faithful

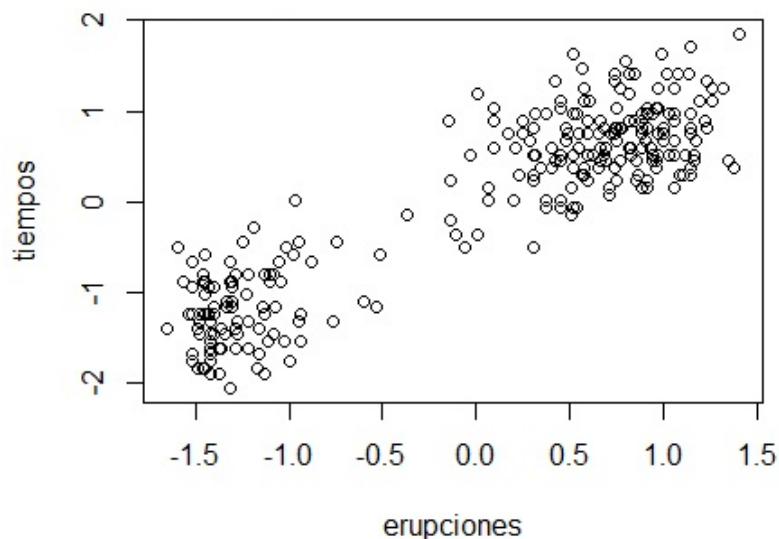


Old Faithful

Datos de Faithful

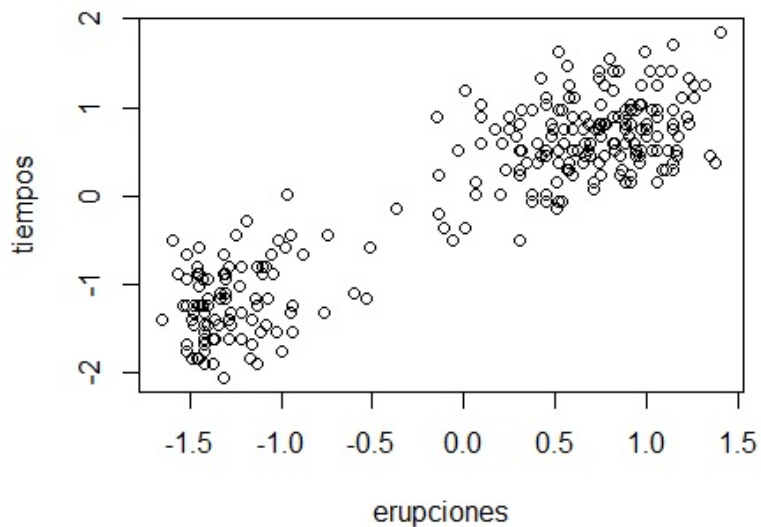


Datos de Faithful (transformados)

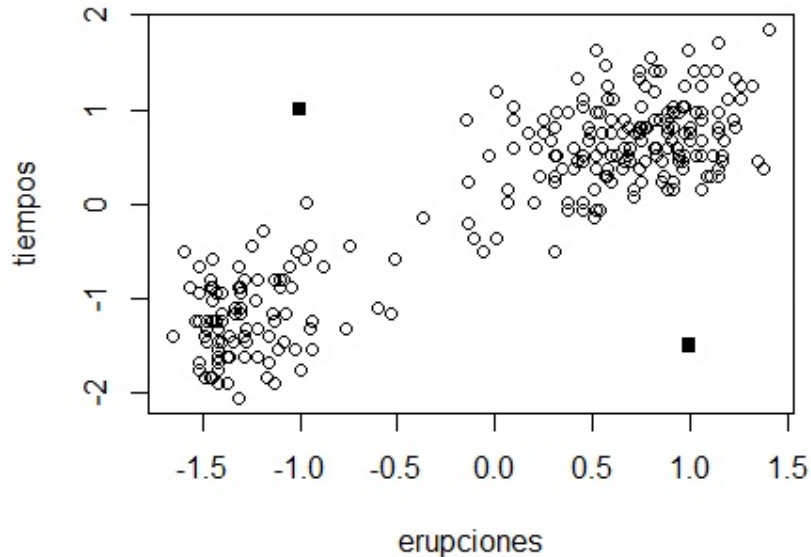


Old Faithful

Datos de Faithful (transformados)

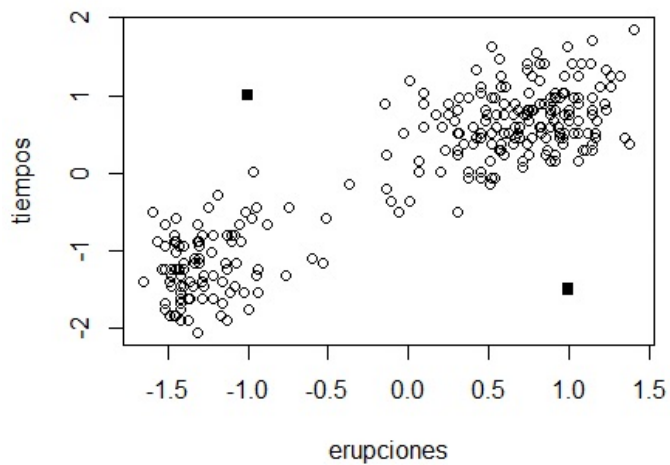


Paso 1



Old Faithful

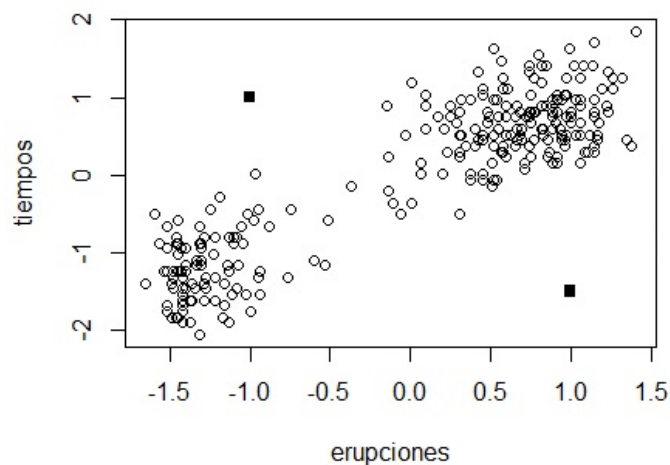
Paso 1



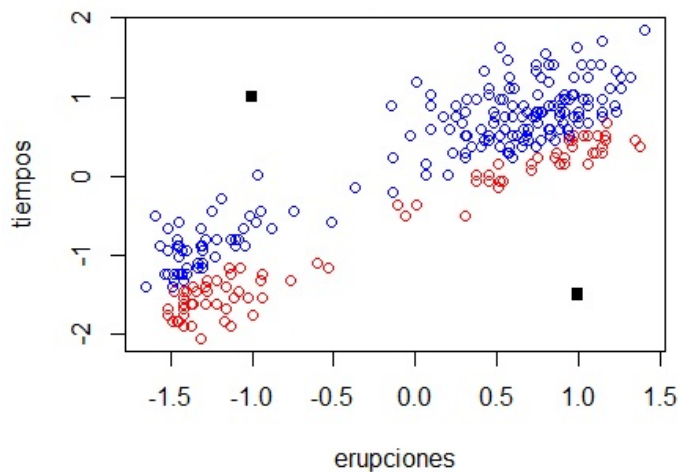
Inicialización

Old Faithful

Paso 1



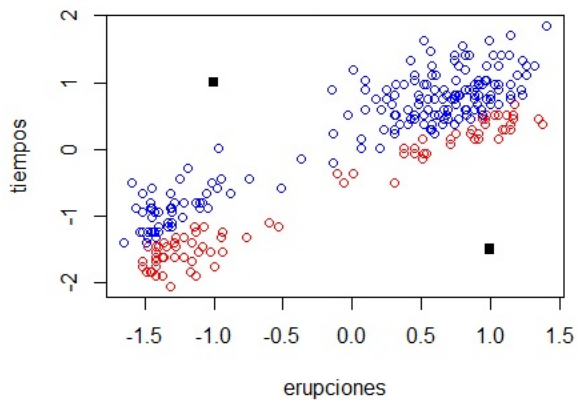
Paso 2



Asignación

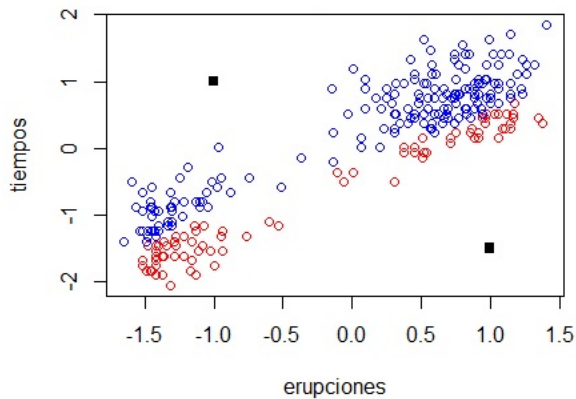
Old Faithful

Paso 2

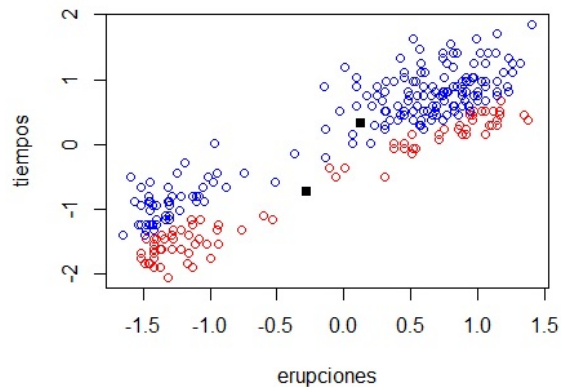


Old Faithful

Paso 2



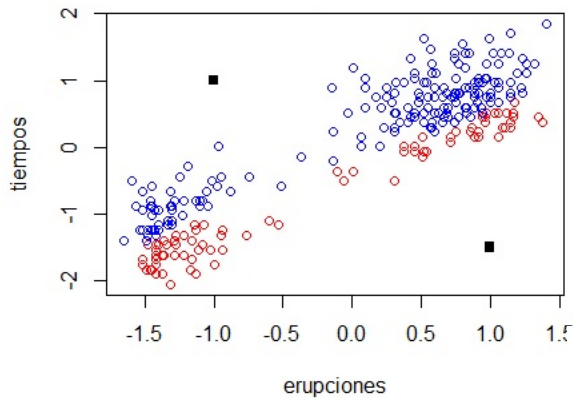
Paso 3



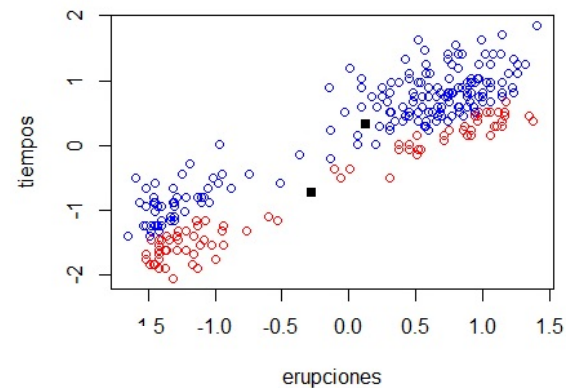
Actualización

Old Faithful

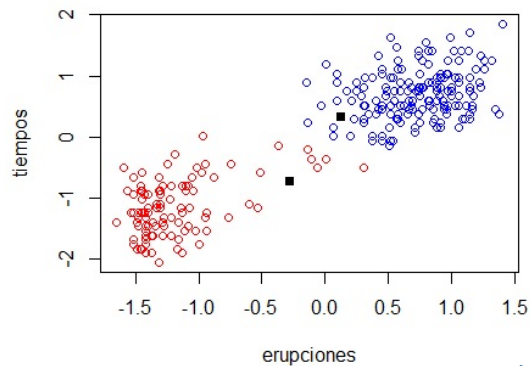
Paso 2



Paso 3



Paso 4

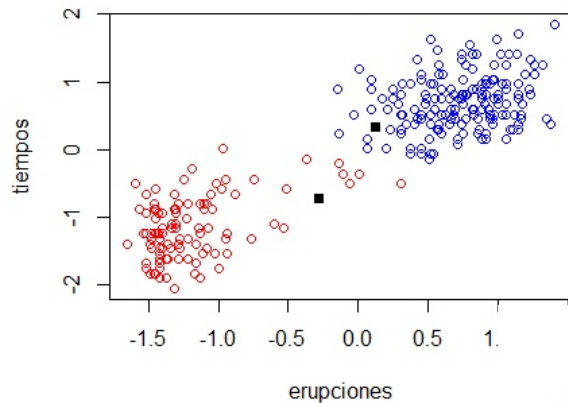


Asignación

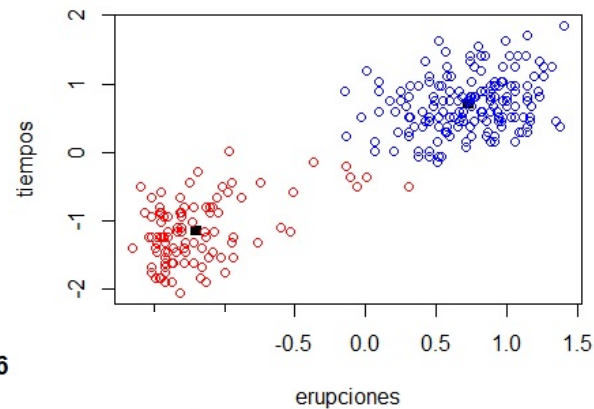
Old Faithful

Actualización

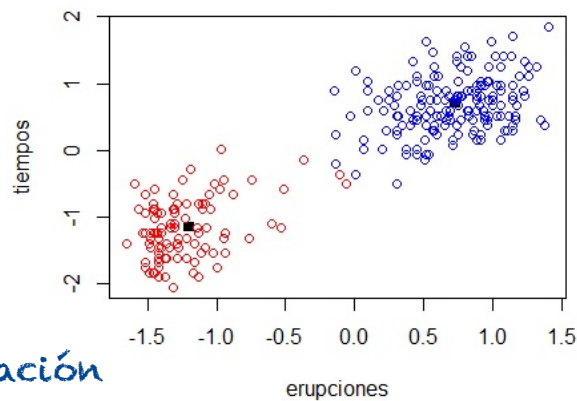
Paso 4



Paso 5



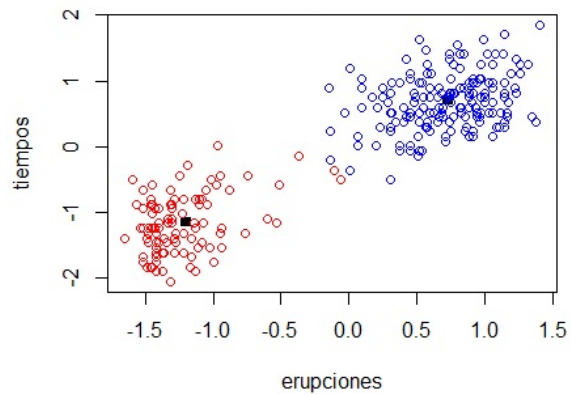
Paso 6



Asignación

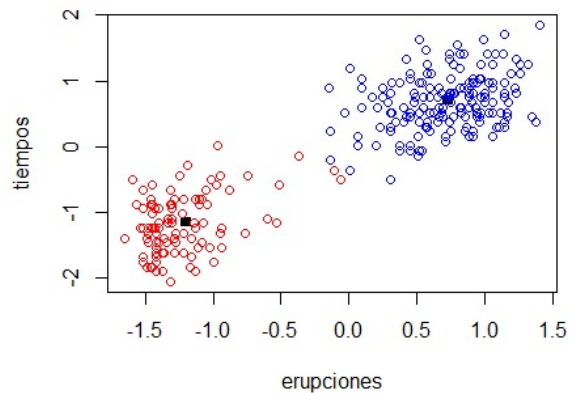
Old Faithful

Paso 6



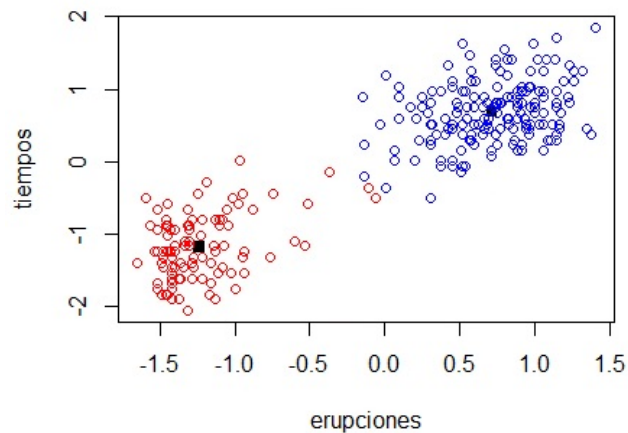
Old Faithful

Paso 6



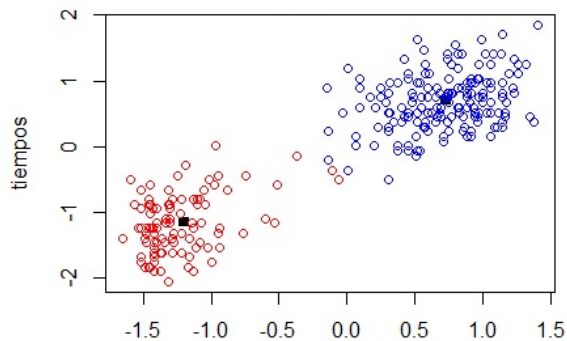
Actualización

Paso 7

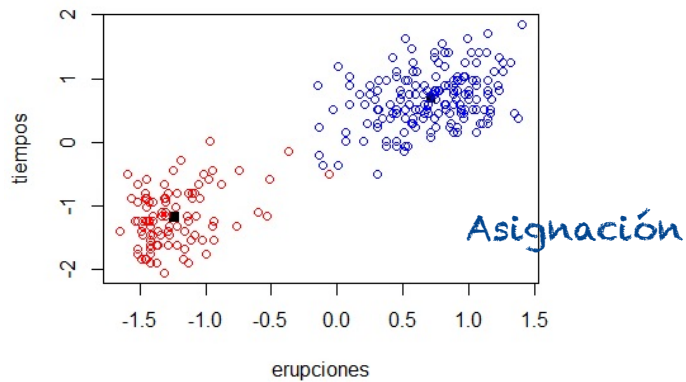


Old Faithful

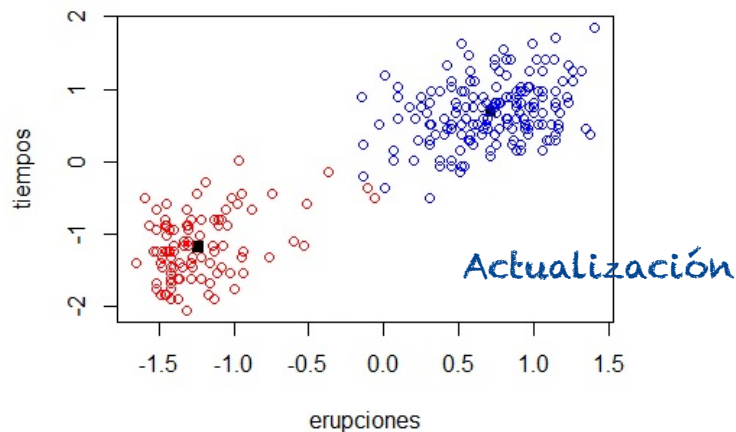
Paso 6



Paso 8

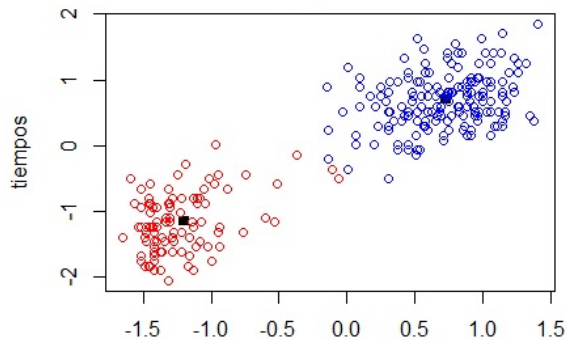


Paso 7

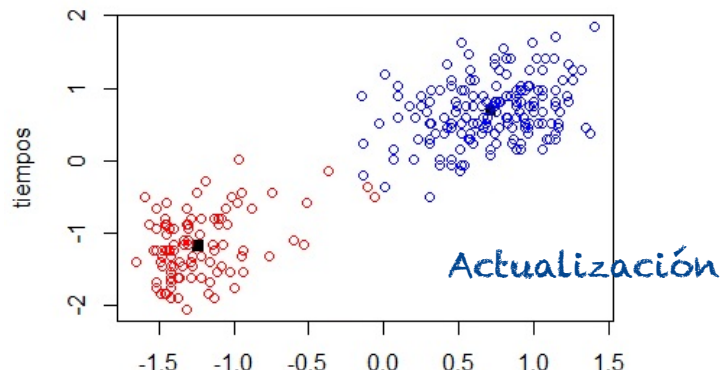


Old Faithful

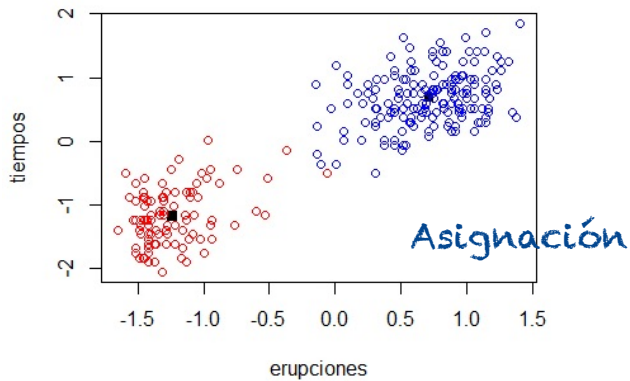
Paso 6



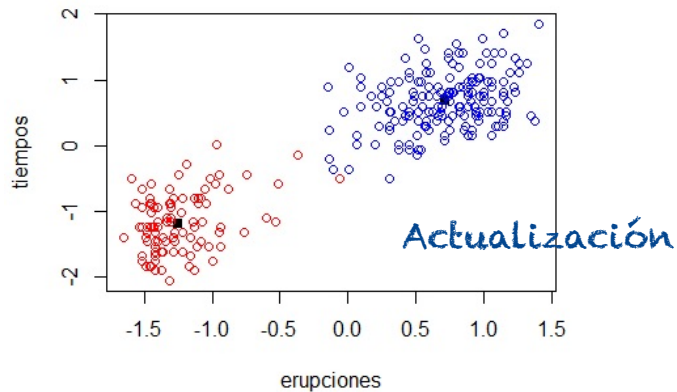
Paso 7



Paso 8



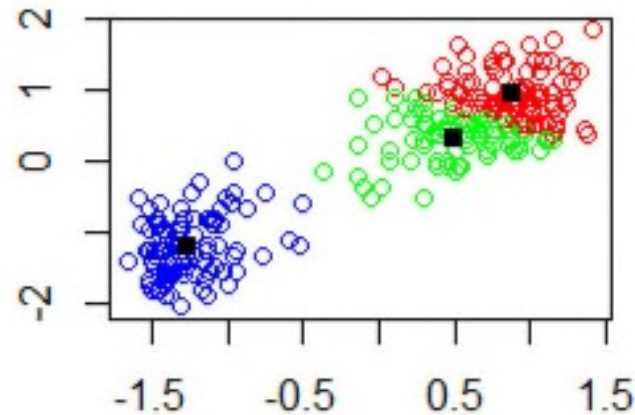
Paso 9



Qué hago si no conozco el número de clusters?

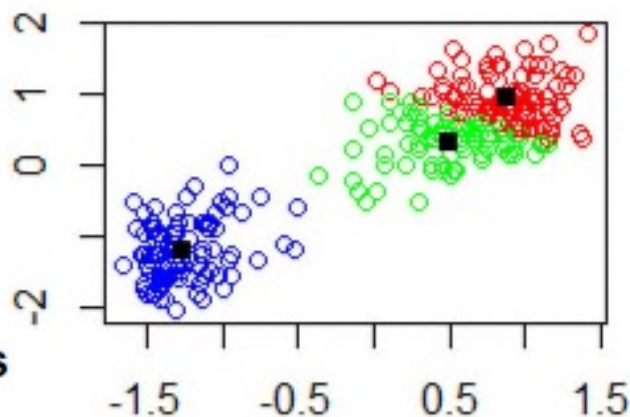
Qué hago si no conozco el número de clusters?

Cluster con 3 grupos

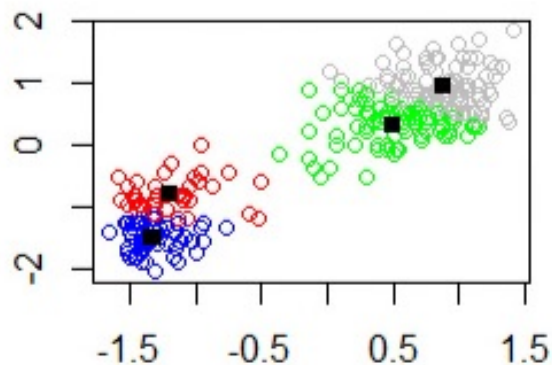


Qué hago si no conozco el número de clusters?

Cluster con 3 grupos



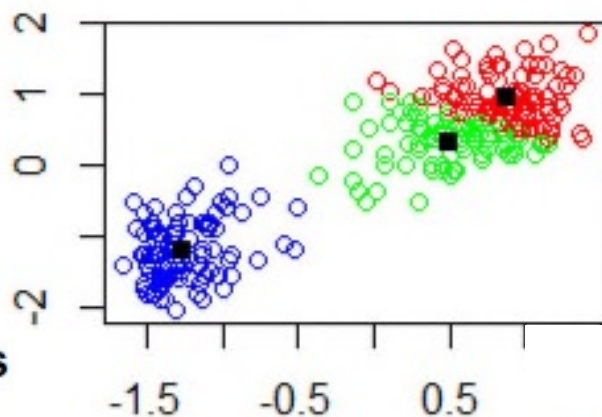
Cluster con 4 grupos



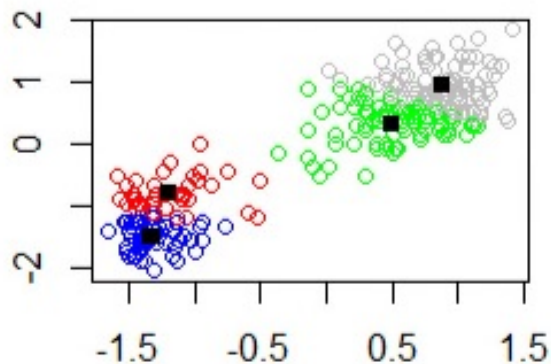
erupciones

Qué hago si no conozco el número de clusters?

Cluster con 3 grupos

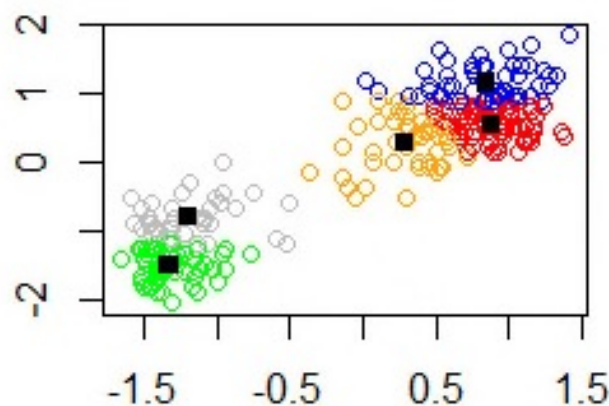


Cluster con 4 grupos



erupciones

Cluster con 5 grupos



Qué hago si no conozco el número de clusters?

Hay formas de medir cuán buena es una clusterización

Qué hago si no conozco el número de clusters?

Hay formas de medir cuán buena es una clusterización

Es importante tener en cuenta que no siempre puedo dibujar

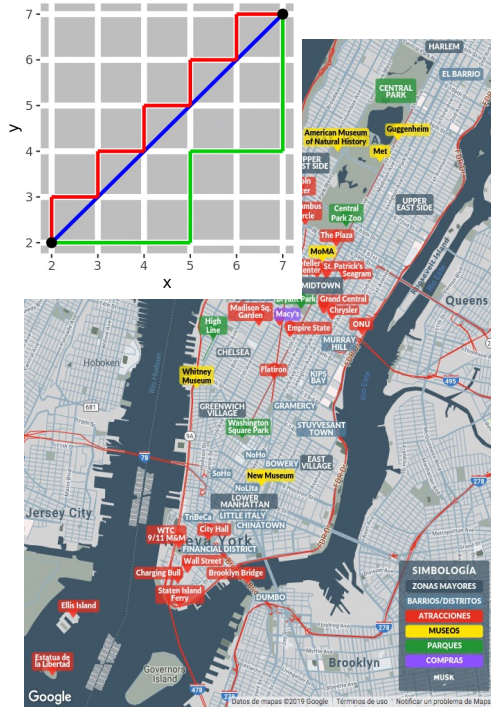
Y que hago con las películas, las imágenes, el volcán y el cajón?



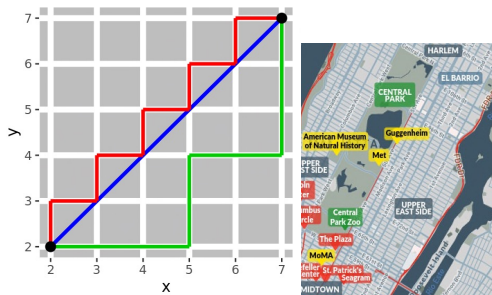
Distancias entre objetos

Distancias entre objetos

Distancias entre objetos

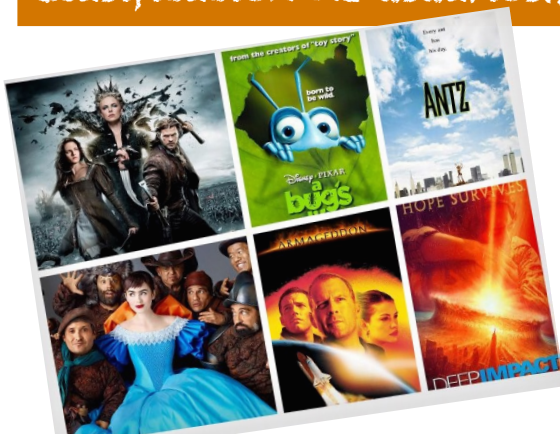


Distancias entre objetos

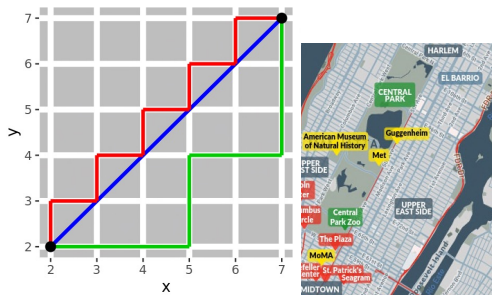


Contenido, resumen, palabras?

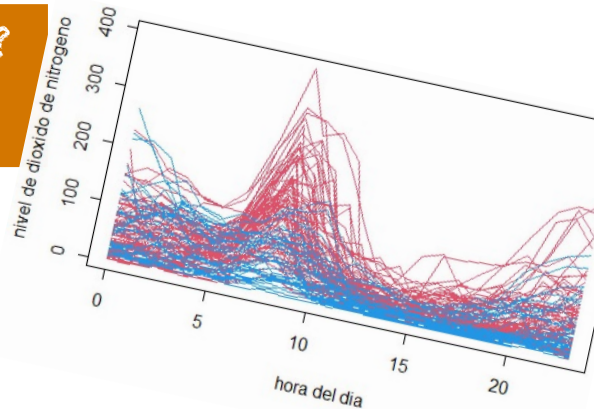
Clasificación de usuarios.?



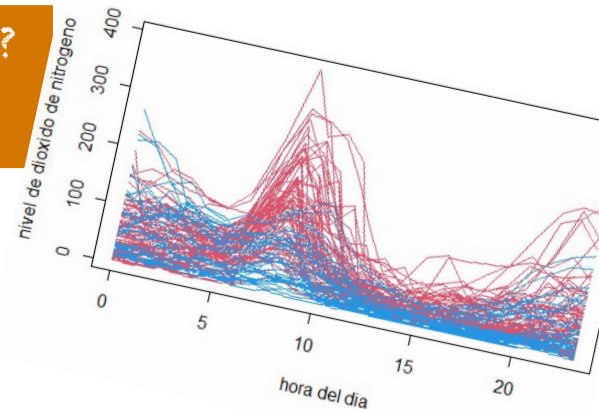
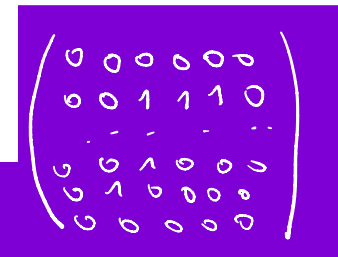
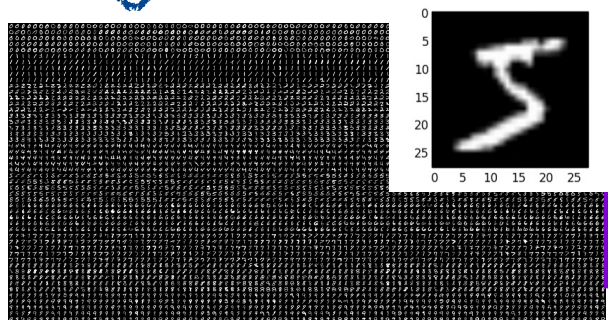
Distancias entre objetos



Contenido, resumen, palabras?
Clasificación de usuarios?



$$\int (f(x) - g(x))^2 dx$$



$$\int (f(x) - g(x))^2 dx$$

Y ahora ya está todo resuelto entonces?

Y ahora ya está todo resuelto entonces?



Y ahora ya está todo resuelto entonces?





Como se escribe el algoritmo?

Como se compara si es mejor 2 o 3 clustres?

Y si no funciona que hago?

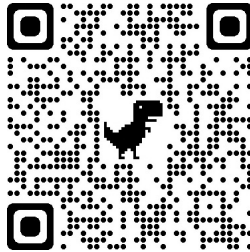
Como se si no funciona?

El algoritmo termina?

Como defino distancias entre películas?

Referencias:

- Página de Walter Sosa Escudero. <https://waltersosa.weebly.com>
- A simple Machine Learning Approach to Geology.
<https://kasumisanchika.medium.com>
- Si están interesados, aquí encontrarán los slides de la charla



Muchas gracias!