

# DESCUBRIENDO PATRONES OCULTOS:

## Aplicación de Biclustering en AtlantoCracies, una Base de Datos NoSQL de la nobleza española en América (Siglos XVI-XIX)

Manuel Díaz-Ordóñez  
Universidad de Sevilla  
[mdiazord@us.es](mailto:mdiazord@us.es)  
X: @meollar

Domingo S. Rodríguez-Baena  
Universidad Pablo de Olavide  
[dsrodbae@upo.es](mailto:dsrodbae@upo.es)

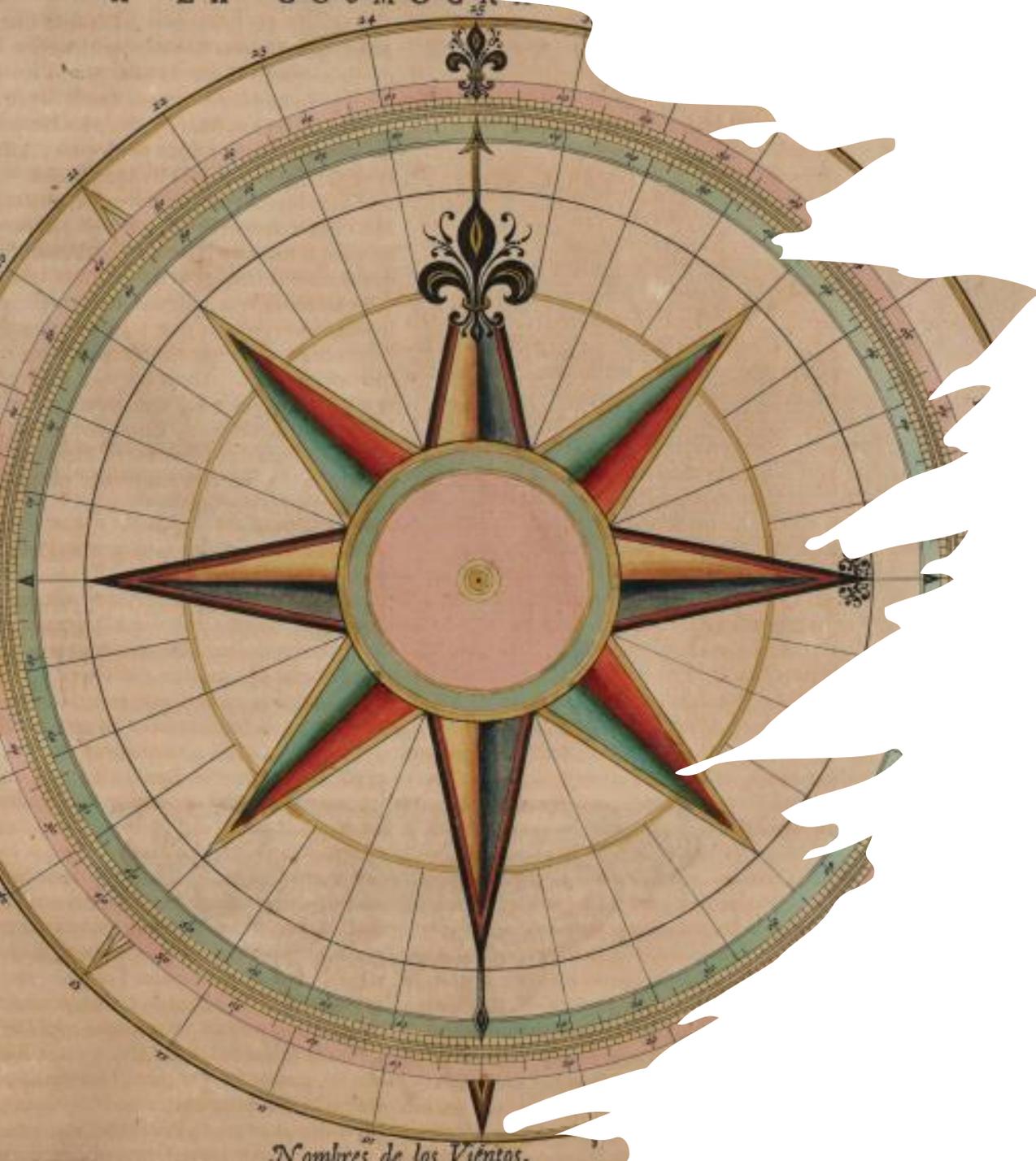
Sevilla, 29 de febrero de 2024

SEMINARIO PERMANENTE

**LOS MUNDOS IBÉRICOS  
Y LA GLOBALIZACIÓN TEMPRANA**

Permanent Seminar: **“Iberian Worlds and Early Globalization”**





# ÍNDICE

1. Introducción
2. Base de Datos AtlantoCracies
3. Biclustering
4. Descubriendo patrones ocultos

# 1. INTRODUCCIÓN

## BASES DE DATOS

DATOS ESTRUCTURADOS

VS

DATOS SEMIESTRUCTURADOS



# INVESTIGACIÓN HISTORIA Y MODELOS SEMIESTRUCTURADOS

## ¿CUÁNDO ELEGIR?

¿La fuente está detenida en el tiempo?

## FUENTES MUY DIVERSAS

- Formatos/objetivos
- Historia del documento
- Nuevas perspectivas

MODIFICACIONES/CAMBIOS  
(corto/medio plazo)

SISTEMA MULTIOBJETIVO  
Más adecuada al desarrollo de la investigación

# 1. INTRODUCCIÓN

## UNA VIDA ¿ES ESTRUCTURADA?

### VIDA

- Fechas
- Lugares

### PROFESIONES

- Fechas
- Lugares
- Tipos



Fte: Banco de Imágenes del Museo Nacional del Prado

### TÍTULOS

- Fechas
- Lugares
- Tipos

### EVENTOS

- Fechas
- Lugares
- Tipos

### RELACIONES

- Personas
- Fechas
- Lugares
- Tipos

## 2. BASE DE DATOS ATLANTOCRACIES



Fte: Biblioteca Nacional de España

PROYECTO FEDER

En busca de las  
aristocracias  
atlánticas.

América y las  
élites en la España  
peninsular, 1492-  
1824.

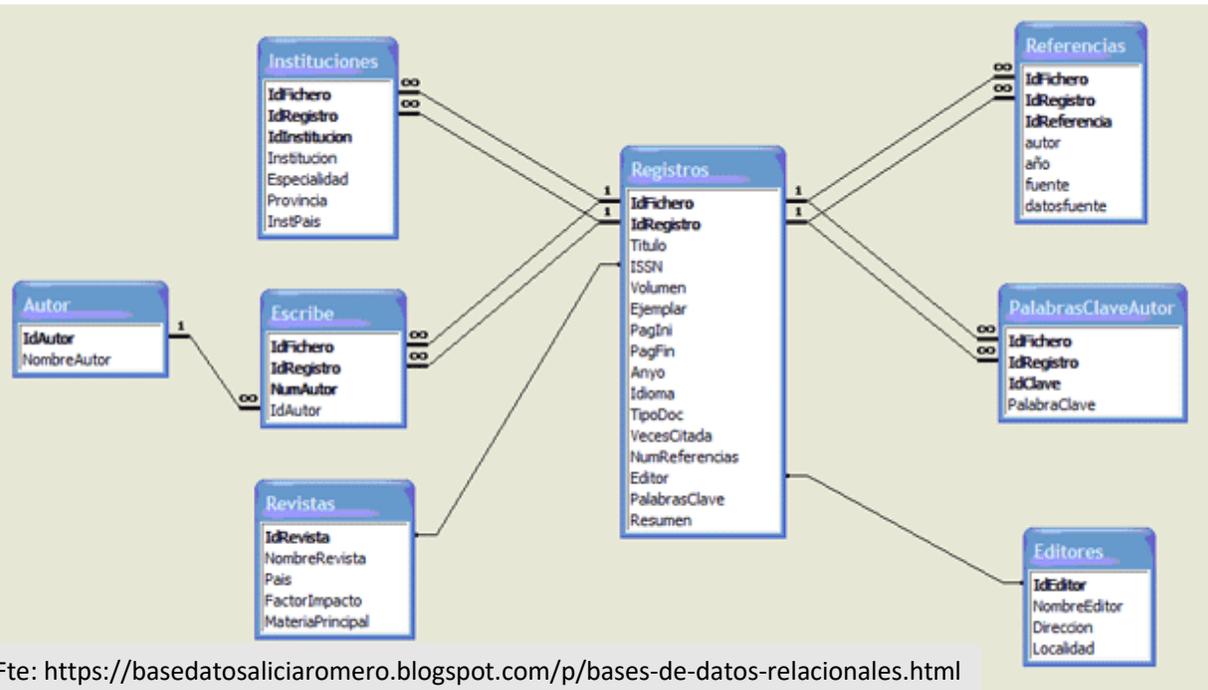
IP B. Yun Casalilla

UPO

# 2. BASE DE DATOS ATLANTOCRACIES

## FLEXIBILIDAD vs RIGIDEZ

## APARIENCIA: LENGUAJE NATURAL



```
_id: ObjectId('609cd66b9aab3a99799344bc')
datebirth: "1716-01-21"
placebirth: "Ciudad de México"
pbirthUTM: "19.419444/-99.145556"
countryDeath: "México"
datedeath: "1786-08-29"
gender: "H"
histDeath: "Virreinato de Nueva España"
name: "Francisco Javier de Aristoarena de Lanz"
pDeathUTM: "25.816667/-100.133333"
pageFr: "Null"
personId: 117
placeDeath: "Zacatecas"
source: "Ramón Maruri Villanueva Personal Database (American Spanish Nobility) ..."
sourceType: "Database"
titles: Array
  0: Object
    continental: "América"
    countryTit: "España"
    dateTit: "1777"
    histTit: "Virreinato de Nueva España"
    nomTit: "Condado de Casa Fiel"
relations: Array
  0: Object
    idPerson: 710
    namePerson: "Josefa Tagle Bracho"
    placeRel: "Ciudad de México"
    placeRUTM: "19.419444/-99.145556"
```



Fte: Banco de Imágenes del Museo Nacional del Prado

Fte: <https://basedatosaliciaromero.blogspot.com/p/bases-de-datos-relacionales.html>

## 2. BASE DE DATOS ATLANTOCRACIES

### EN NÚMEROS

- **Personas**

The screenshot shows the MongoDB Atlas interface for the 'AtlantoCracies.Nobility' database. The top navigation bar includes 'Documents', 'Aggregations', 'Schema', 'Explain Plan', 'Indexes', and 'Validation'. The 'Aggregations' tab is active. The top right corner displays statistics: DOCUMENTS 3.7k, STORAGE SIZE 1.7MB, AVG. SIZE 1.3KB, INDEXES 2, TOTAL SIZE 245.8KB, and AVG. SIZE 122.9KB. Below the navigation bar, there are controls for 'COLLATION', 'Untitled', 'SAVE', 'SAMPLE MODE', and 'AUTO PREVIEW'. The main content area shows '3708 Documents in the Collection' and a 'Preview of Documents in the Collection'. Two document snippets are visible, each with a unique ObjectId and various fields like continentBirth, countryBirth, datebirth, gender, hist30Birth, histBirth, and name.

- Caballeros (Santiago, Calatrava, Alcántara, Carlos III, Malta)
- Padre
- Madre
- Hermanos
- Testigos de prueba (**Innovación**)

## 2. BASE DE DATOS ATLANTOCRACIES

### EN NÚMEROS

- Personas
- Lugares

- 1.481 Lugares (localidades, edificios, batallas, etc.)
- 42 Estados históricos (reinos, virreinos, capitanías, gobernaciones, etc.)

## 2. BASE DE DATOS ATLANTOCRACIES

### EN NÚMEROS

- Personas
- Lugares
- Eventos (relaciones)

- 104 eventos (matrimonio, duelo, guerra, enfermedad, estudió, etc.)

## 2. BASE DE DATOS ATLANTOCRACIES

### EN NÚMEROS

- Personas
- Lugares
- Eventos/relaciones
- **Títulos**

- 606 títulos nobiliarios (condes, marqueses, ducados, etc.)

## 2. BASE DE DATOS ATLANTOCRACIES

### EN NÚMEROS

- Personas
- Lugares
- Eventos/relaciones
- Títulos
- Tipos de relación

- 21 tipos de relación (parentesco, **testigos**, notarial, matrimonio, amistad, socio, subordinado, etc.)

## 2. BASE DE DATOS ATLANTOCRACIES

### EN NÚMEROS

- Personas
- Lugares
- Eventos/relaciones
- Títulos
- Tipos de relación
- Fuentes históricas

- 18 fuentes (Archivos, bibliografía, bases de datos, etc.)

## 2. BASE DE DATOS ATLANTOCRACIES

### EN NÚMEROS

- Personas
- Lugares
- Eventos/relaciones
- Títulos
- Tipos de relación
- Fuentes históricas
- Posiciones/cargos

- 1054 (religiosos, políticos, judiciales, militares, educación, comerciantes, médicos, etc.)

```
/* 13 */
{
  "_id" : ObjectId("609cd6fd9aab3a997993eafb"),
  "continentBirth" : "América",
  "countryBirth" : "Cuba",
  "datebirth" : "1720-04-30",
  "datedeath" : "1781",
  "gender" : "H",
  "hist30Birth" : "Virreinato de Nueva España",
  "histBirth" : "Virreinato de Nueva España",
  "name" : "Pedro José Calvo de la Puerta y Arango",
  "pageFr" : "Null",
  "pbirthUTM" : "23.40463807274443/-82.19509495405939",
  "personId" : 187,
  "placebirth" : "La Habana",
  "source" : "Guillermo Lohmann Villena Los americanos en las órdenes nobilarias Madrid CSIC 1993",
  "sourceType" : "Book",
  "titles" : [
    {
      "continental" : "América",
      "countryTit" : "España",
      "dateTit" : "1766",
      "numTit" : 1,
      "benefactor" : 187,
      "histTit" : "Virreinato de Nueva España",
      "nomTit" : "Condado de Buena Vista"
    }
  ],
  "relations" : [
    {
      "idPerson" : 788,
      "namePerson" : "Catalina Josefa O'Farril",
      "placeRel" : "La Habana",
      "placeRUTM" : "23.40463807274443/-82.19509495405939",
      "treatment" : "Doña",
      "typeRel" : "Matrimonio",
      "dateRelF" : "1746-12-06"
    }
  ],
}
```

```
"events" : [  
  {  
    "dateEv" : "1720-04-30",  
    "placeEv" : "Catedral de la Virgen María de la Concepción Inmaculada (Antigua parroquia de San Cristóbal/La Habana)",  
    "placeEvUTM" : "23.141186686101957/-82.35171369152641",  
    "typeEv" : "Bautizo"  
  },  
  {  
    "dateEv" : "1746-12-06",  
    "placeEv" : "Catedral de la Virgen María de la Concepción Inmaculada (Antigua parroquia de San Cristóbal/La Habana)",  
    "placeEvUTM" : "23.141186686101957/-82.35171369152641",  
    "typeEv" : "Boda"  
  }  
],  
"positions" : [  
  {  
    "countryPos" : "Cuba",  
    "namePos" : "Regidor perpetuo",  
    "placePos" : "La Habana",  
    "placePosUTM" : "23.40463807274443/-82.19509495405939"  
  }  
],  
"treatment" : "Don"  
}
```

## 2. BASE DE DATOS ATLANTOCRACIES

DIFUSIÓN:

<https://atlantocracies.com/>



# 2. BASE DE DATOS ATLANTOCRACIES

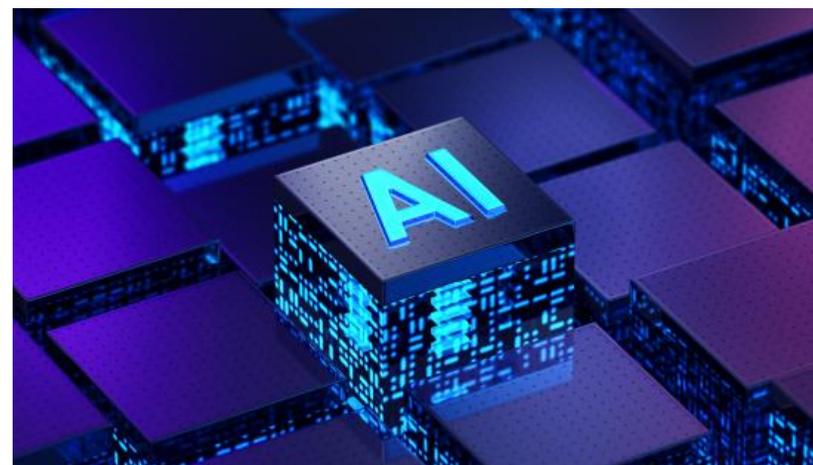
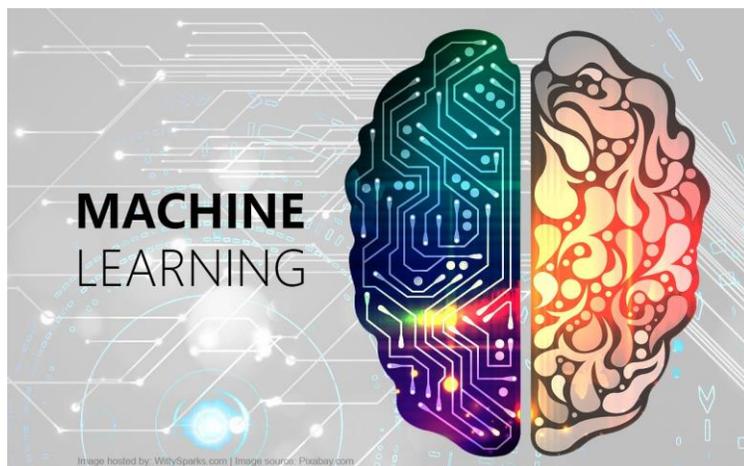
DIFUSIÓN:

<https://atlantocracies.com/>



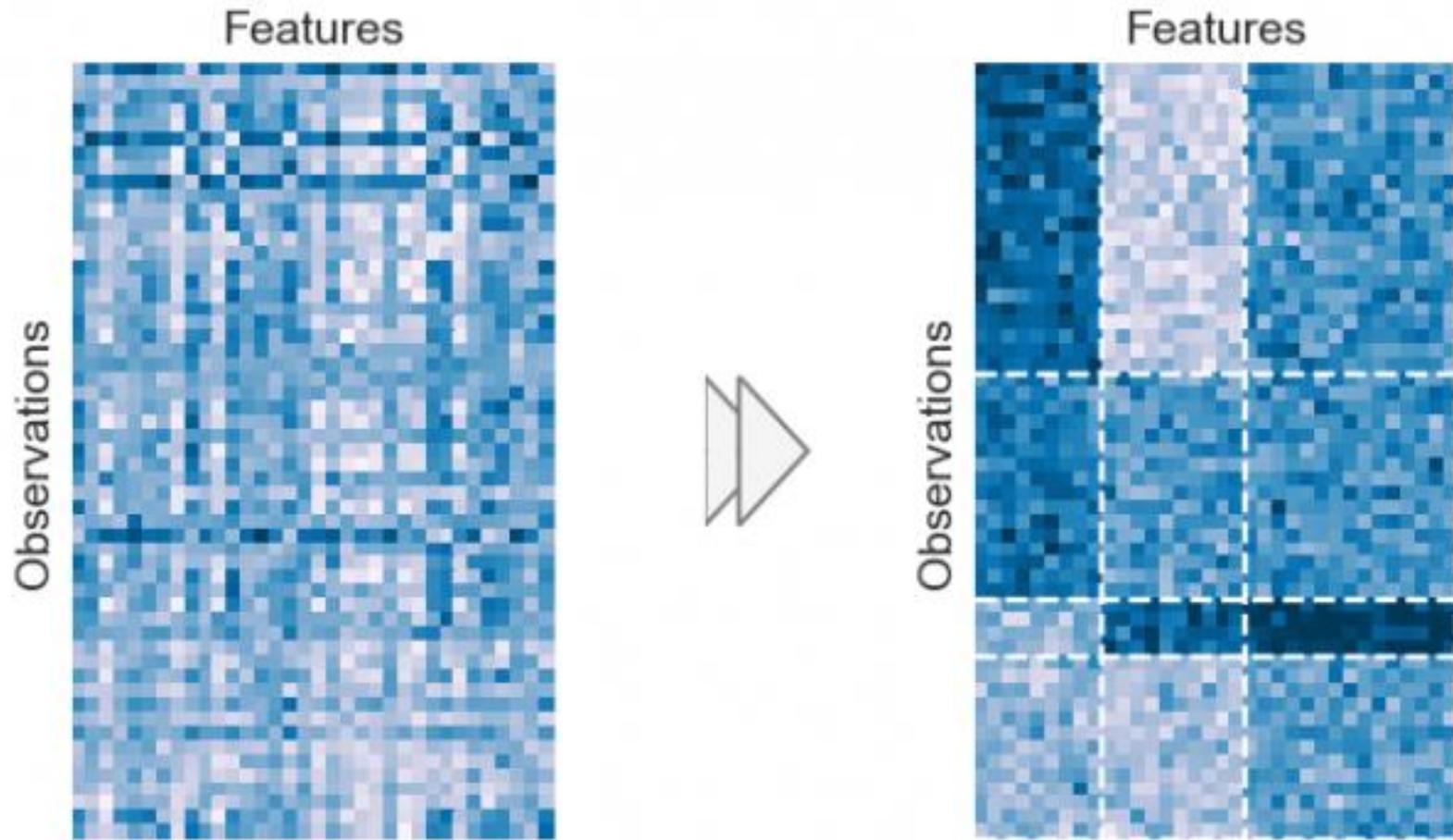
# 3. BICLUSTERING

- DEFINICIÓN → Técnica de aprendizaje no supervisado (Machine Learning) que agrupa elementos con características comunes teniendo en cuenta simultáneamente dos dimensiones.



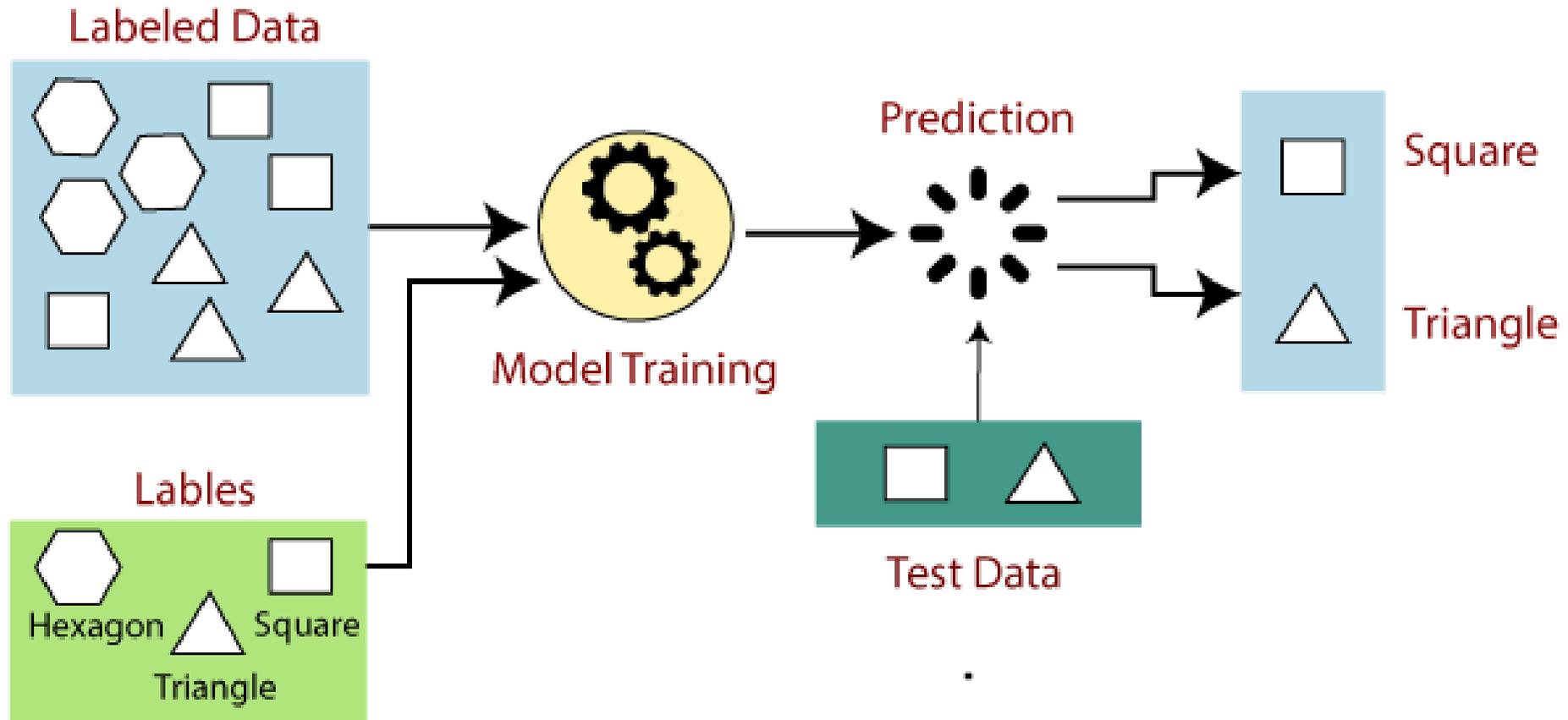
# 3. BICLUSTERING

## *UNSUPERVISED LEARNING*



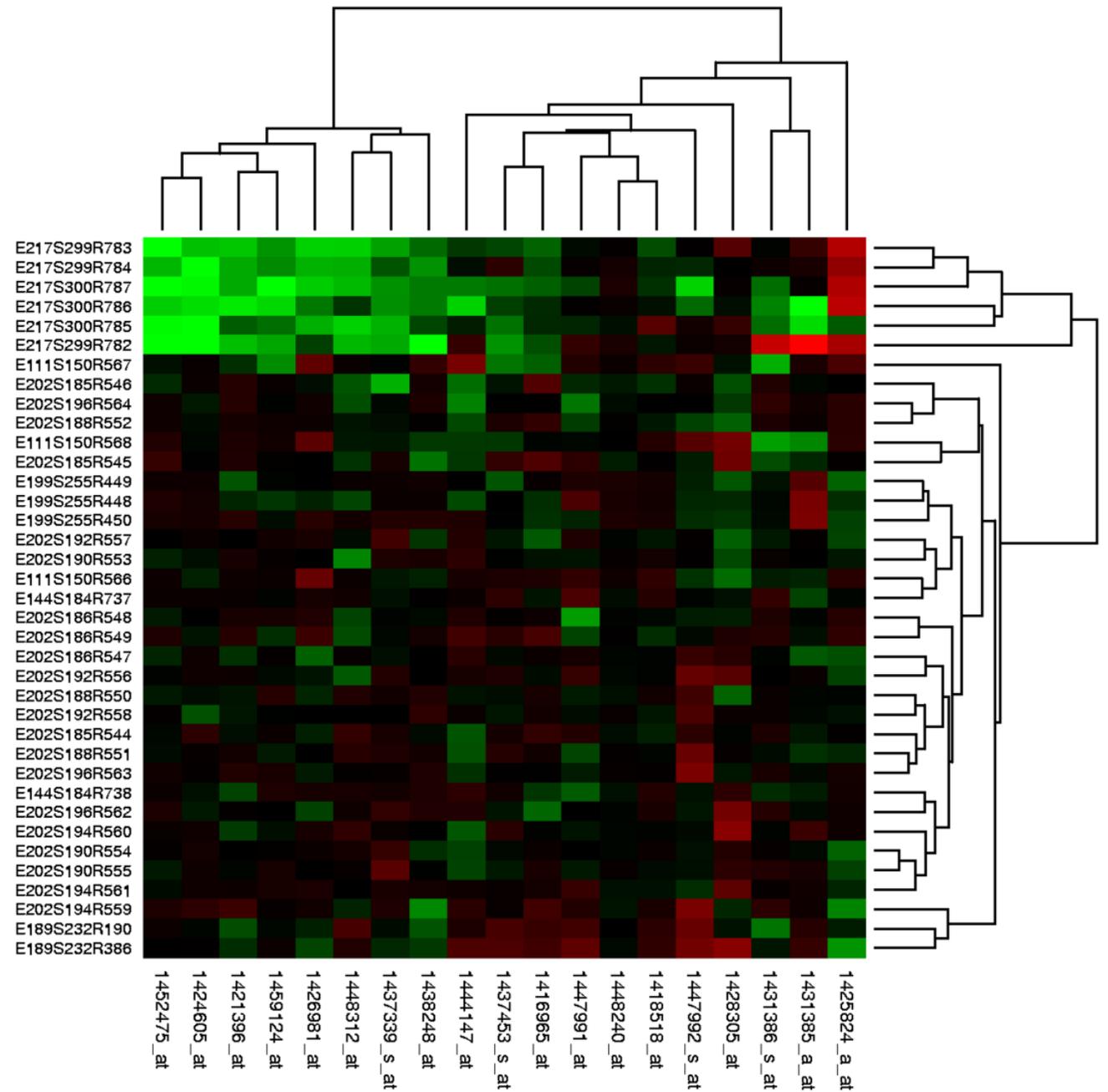
# 3. BICLUSTERING

## *SUPERVISED LEARNING*



# 3. BICLUSTERING

- BIOINFORMÁTICA.
- ANÁLISIS DE DATOS DE EXPRESIÓN GENÉTICA



# 3. BICLUSTERING

- ANTECEDENTE DEL BICLUSTERING → EL CLUSTERING

	A	B	C	D	E	F	G	H
Gen 1	Orange	Yellow	White	Orange	Yellow	Yellow	Orange	Orange
Gen 2	White	White	White	White	White	White	White	White
Gen 3	White	White	White	White	White	White	White	White
Gen 4	Orange	Yellow	White	Orange	Yellow	Yellow	Orange	Orange
Gen 5	White	White	White	White	White	White	White	White
Gen 6	White	Yellow	White	White	Yellow	Yellow	White	White
Gen 7	White	Yellow	White	White	Yellow	Yellow	White	White
Gen 8	White	White	White	White	White	White	White	White
Gen 9	Orange	Yellow	White	Orange	Yellow	Yellow	Orange	Orange

	A	B	C	D	E	F	G	H
Gen 1	Orange	Yellow	White	Orange	Yellow	Yellow	Orange	Orange
Gen 4	Orange	Yellow	White	Orange	Yellow	Yellow	Orange	Orange
Gen 9	Orange	Yellow	White	Orange	Yellow	Yellow	Orange	Orange

# 3. BICLUSTERING

	A	B	C	D	E	F	G	H
Gen 1	Orange	Yellow	White	Orange	Yellow	Yellow	Orange	Orange
Gen 2	White	White	White	White	White	White	White	White
Gen 3	White	White	White	White	White	White	White	White
Gen 4	Orange	Yellow	White	Orange	Yellow	Yellow	Orange	Orange
Gen 5	White	White	White	White	White	White	White	White
Gen 6	White	Yellow	White	White	Yellow	Yellow	White	White
Gen 7	White	Yellow	White	White	Yellow	Yellow	White	White
Gen 8	White	White	White	White	White	White	White	White
Gen 9	Orange	Yellow	White	Orange	Yellow	Yellow	Orange	Orange

	A	B	D	E	F	G	H
Gen 1	Orange	Yellow	Orange	Yellow	Yellow	Orange	Orange
Gen 4	Orange	Yellow	Orange	Yellow	Yellow	Orange	Orange
Gen 9	Orange	Yellow	Orange	Yellow	Yellow	Orange	Orange

BICLUSTER 1

	B	E	F
Gen 1	Yellow	Yellow	Yellow
Gen 4	Yellow	Yellow	Yellow
Gen 6	Yellow	Yellow	Yellow
Gen 7	Yellow	Yellow	Yellow
Gen 9	Yellow	Yellow	Yellow

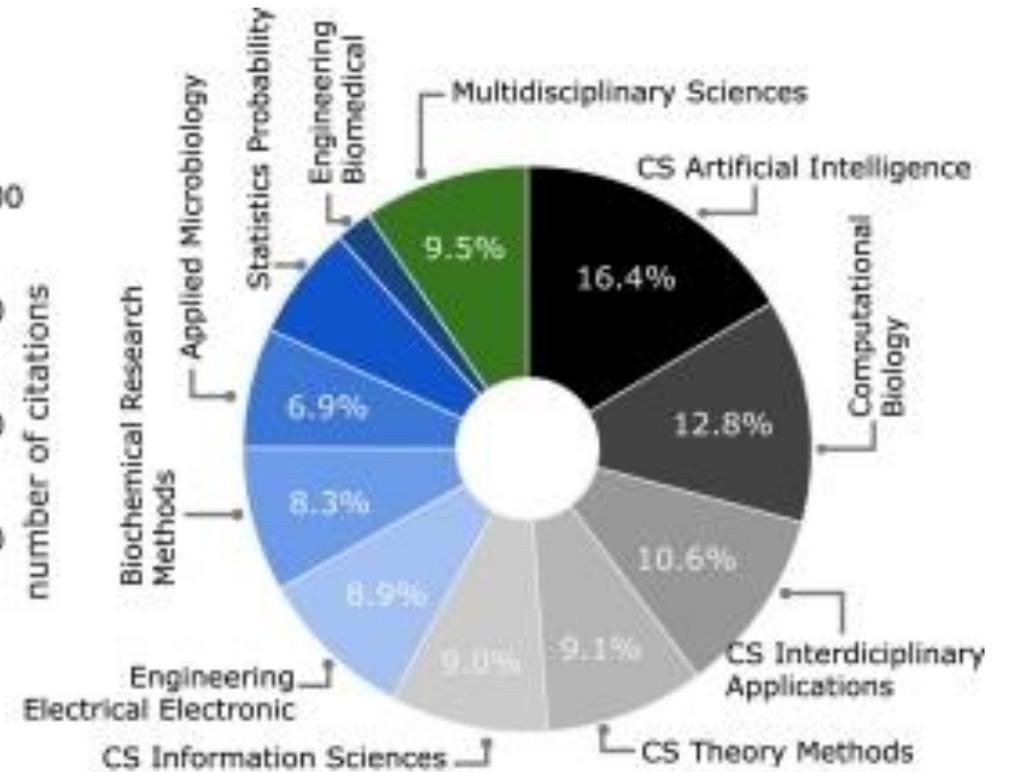
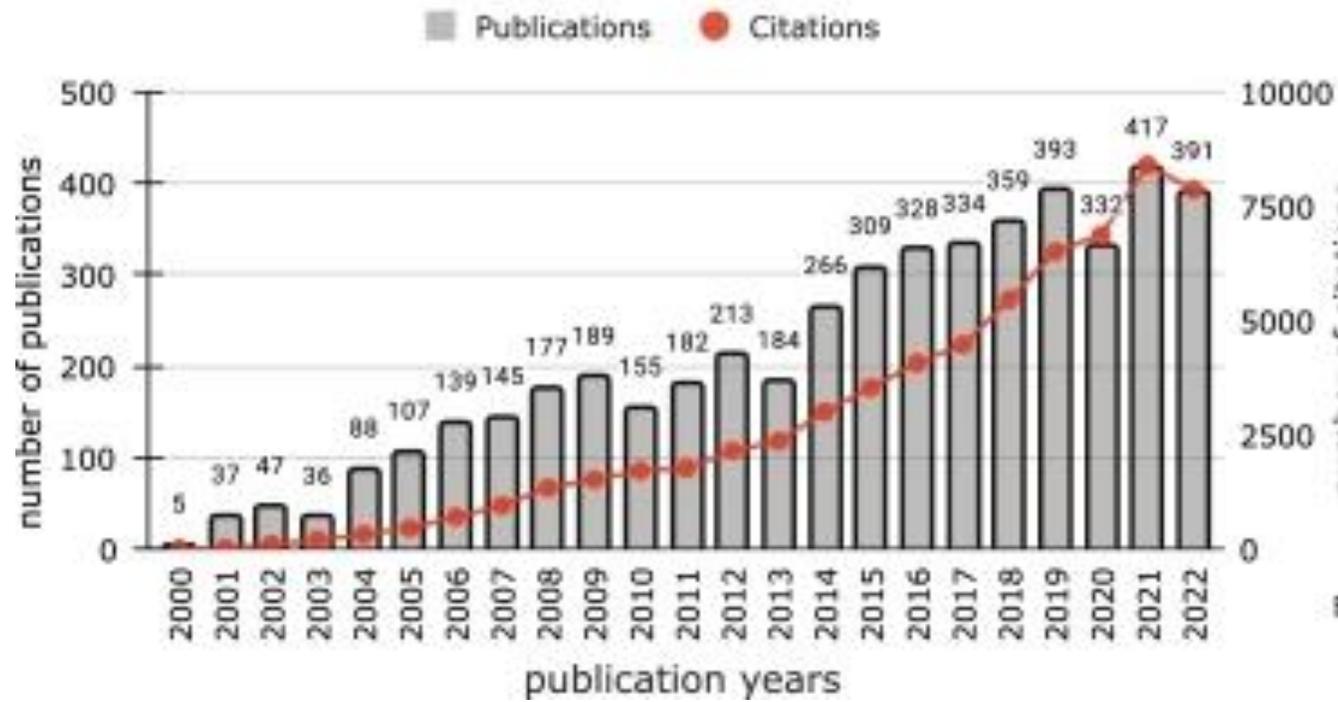
BICLUSTER 2

- Cheng Y, Church GM. Biclustering of expression data. Proc Int Conf Intell Syst Mol Biol. 2000;8:93-103.

### 3. BICLUSTERING

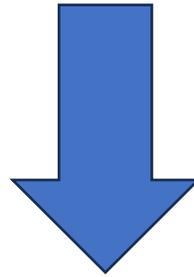
- ✓ Se tienen en cuenta las dos dimensiones simultáneamente.
- ✓ El agrupamiento NO es DISJUNTO.
- ✓ Los biclusters son submatrices dentro de la misma matriz.
- ✓ Se pueden seguir distintos modelos de agrupamiento.
- ✓ Aumentan considerablemente el número de posibles soluciones.
- ✓ Necesidad de una evaluación de resultados.
- ✓ Atención con el nivel de solapamiento.

# 3. BICLUSTERING



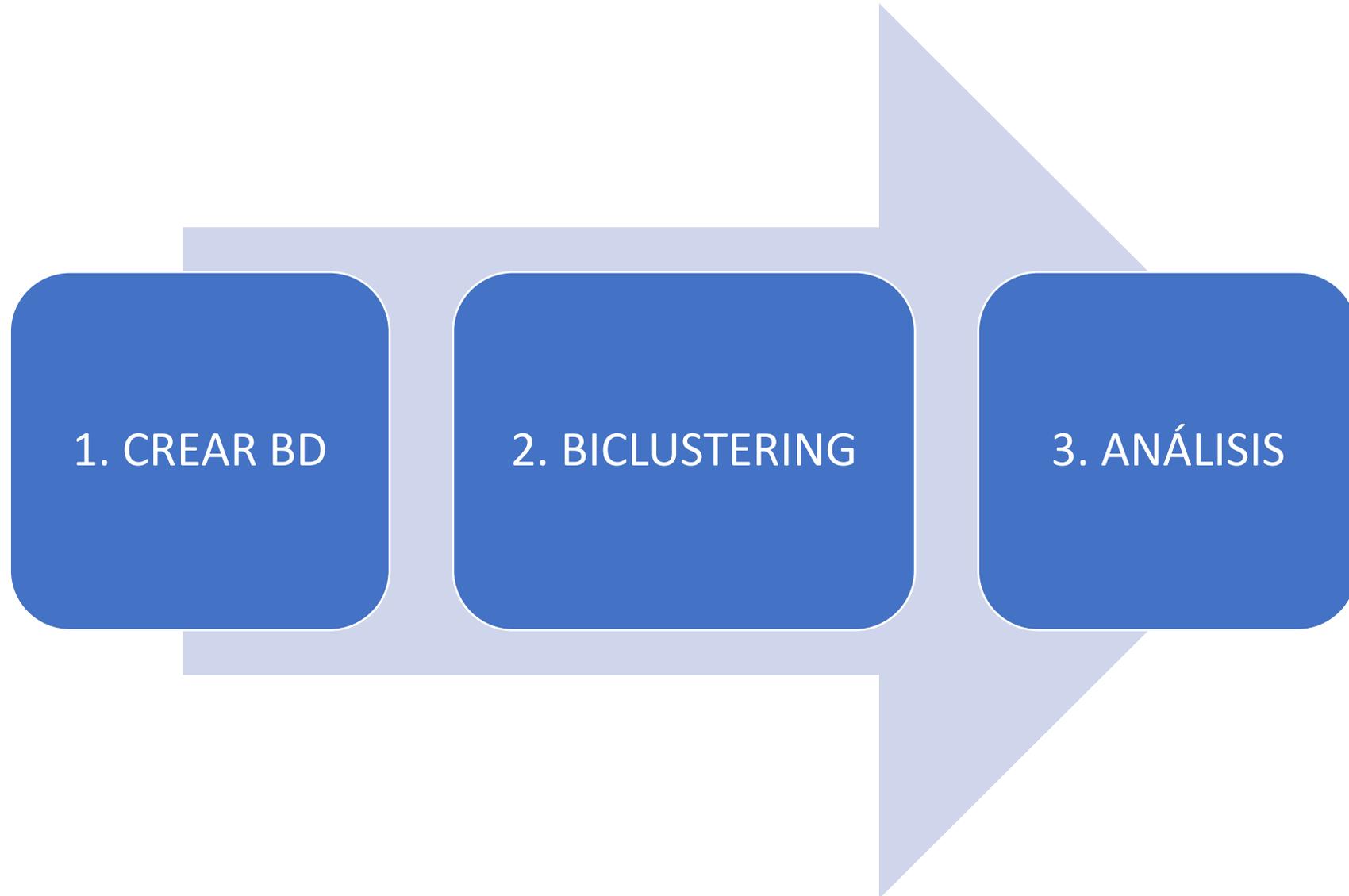
## 4. DESCUBRIENDO PATRONES OCULTOS.

*“Ayudas a Proyectos Interdisciplinarios”* VI Plan Propio de Investigación y Transferencia (2023-2026) de la Universidad Pablo de Olavide

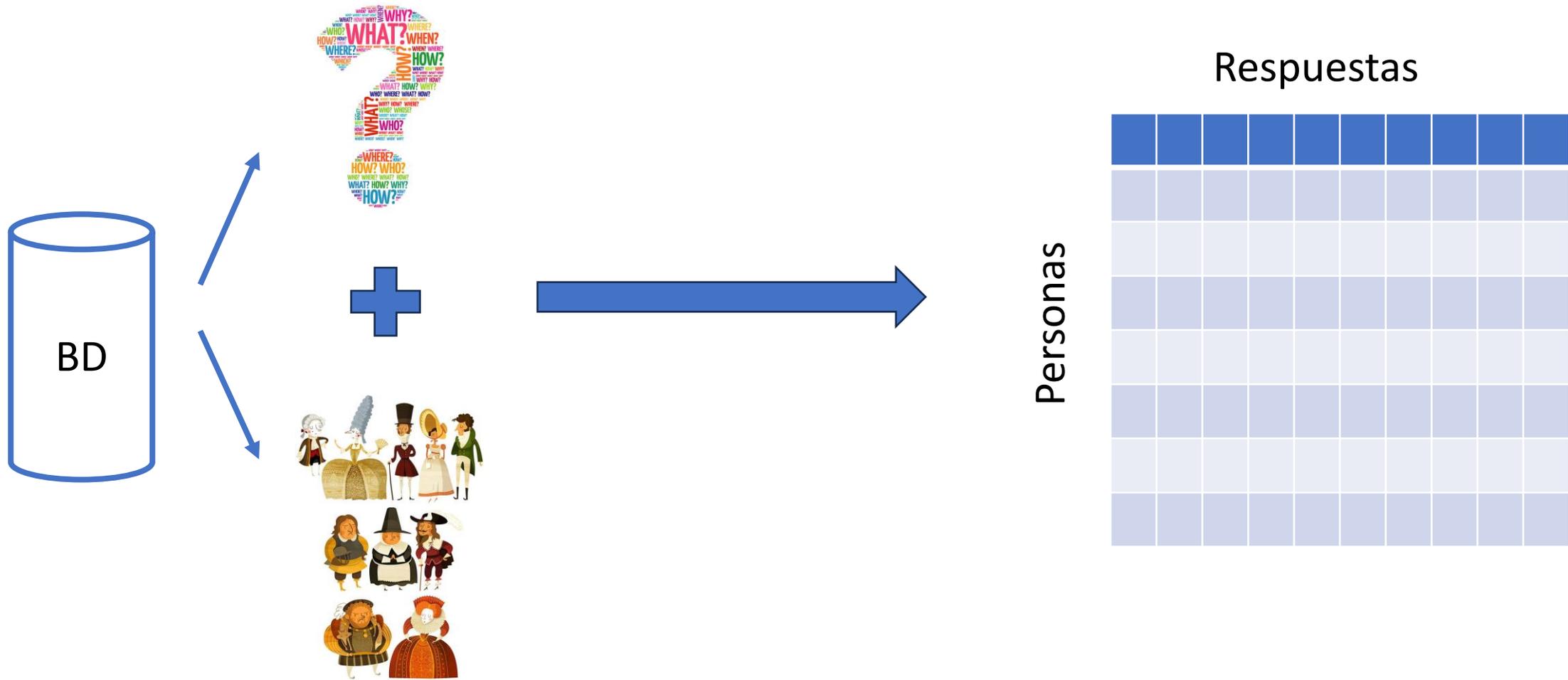


***DESARROLLO Y EXPLOTACIÓN DE BASES DE DATOS NoSQL, PARA EL ESTUDIO DE LAS NOBLEZAS TRANSFRONTERIZAS EN LA ÉPOCA MODERNA. LA BASE DE DATOS ATLANTOCRACIES***

## 4. DESCUBRIENDO PATRONES OCULTOS.



# 4.1 Creación de la BD



### ¿QUE TIPO DE PREGUNTAS?

- Relacionadas con **tiempo**: nacimiento, muerte, obtención título de orden, etc.
- Relacionadas con el **espacio**: lugar nacimiento, muerte, obtención título, lugar matrimonios, etc.
- **Relaciones** con otros **personajes**: matrimonios, informante de título, etc.
- **Características** propias: edad fallecimiento, edad obtención título, número de títulos, ...

### **Primer desafío:**

- Biclustering basado en heurísticas matemáticas.
  
- Respuestas a las preguntas muy heterogéneas:
  - Rango de años
  - Rango de valores numéricos
  - Lista de posibles valores alfanuméricos (combinan letras y números)
  - Binarias: Si / No
  - Numéricas

## 4.2 Biclustering



Necesitamos **pre-procesar** la **matriz** para obtener un formato numérico.



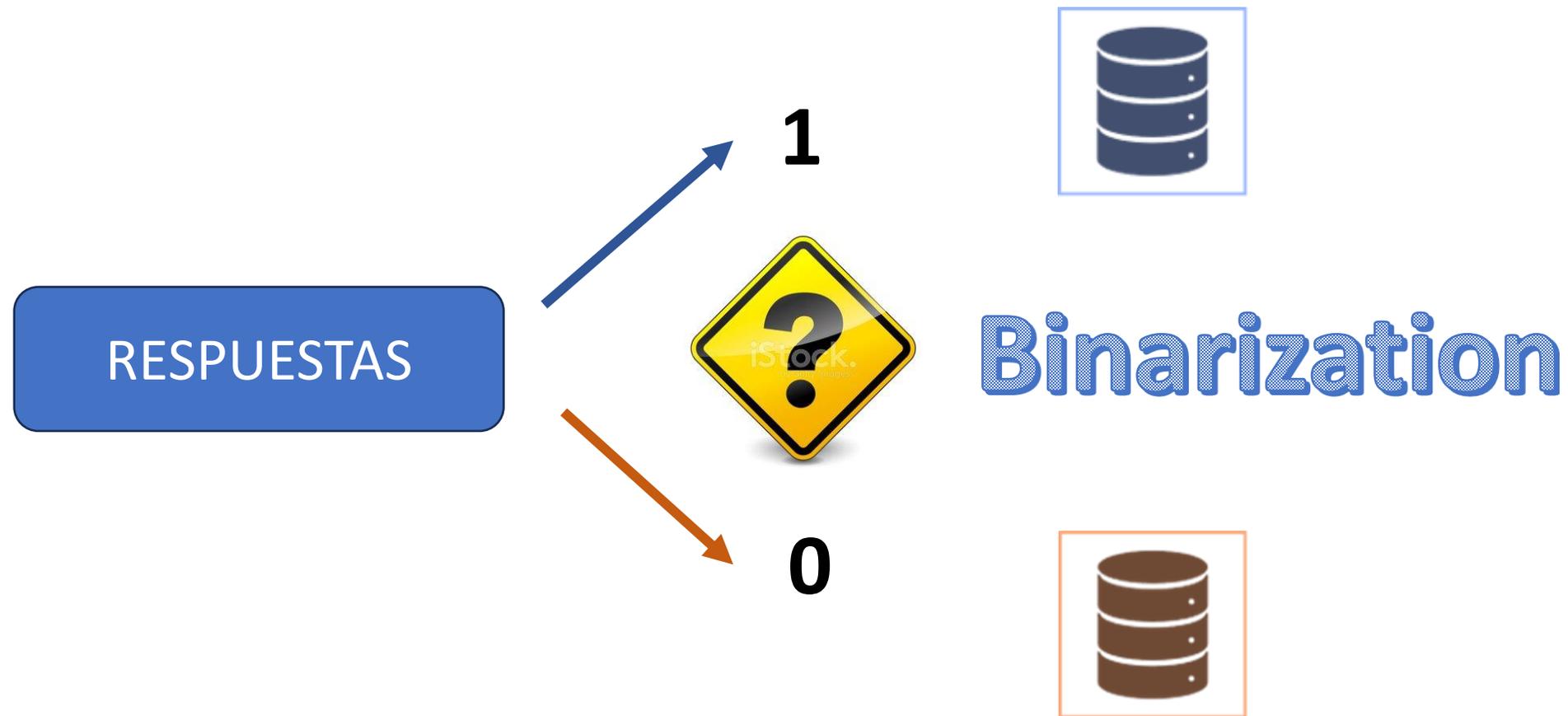
La opción más sencilla: **Bases de datos binarias**

1	1
1	1
1	1
1	1

Los **biclusters** serán submatrices **constantes de 1s**

0	1	0	0	0	1	1	1	0	0
0	0	0	0	1	0	0	1	0	0
1	0	1	0	0	0	0	0	1	0
0	1	0	0	1	1	0	1	0	1
1	1	0	0	1	0	1	0	0	1
0	1	0	0	0	1	0	0	1	0
1	1	0	0	1	0	0	0	1	1
1	1	1	0	0	0	0	1	0	0
1	1	0	1	0	0	0	0	1	1
0	0	0	1	1	1	0	0	1	1

## 4.2 Biclustering



Domingo S. Rodriguez-Baena, Antonio J. Perez-Pulido, Jesus S. Aguilar-Ruiz, **A biclustering algorithm for extracting bit-patterns from binary datasets**, *Bioinformatics*, Volume 27, Issue 19, October 2011, Pages 2738–2745, <https://doi.org/10.1093/bioinformatics/btr464>

# anking

## 4.3 Análisis



- **Bicluster 1**
- **Bicluster 2**
- **Bicluster 3**
- **Bicluster 4**
- ....
- ....
- ....
- ....
- **Bicluster N**

## 4.3 Análisis



MUCHAS GRACIAS POR SU  
ATENCIÓN....¿PREGUNTAS?

Sevilla, 29 de febrero de 2024

Manuel Díaz-Ordóñez y Domingo S. Rodríguez-Baena

