

# Escalamiento Multidimensional (MDS)

José A. Perusquía Cortés



Análisis Multivariado Semestre 2023-2



- **¿De qué va?**

Un conjunto de métodos enfocados en reducir la dimensión usando como criterio preservar la distancia entre observaciones.

- **¿De qué va?**

Un conjunto de métodos enfocados en reducir la dimensión usando como criterio preservar la distancia entre observaciones.

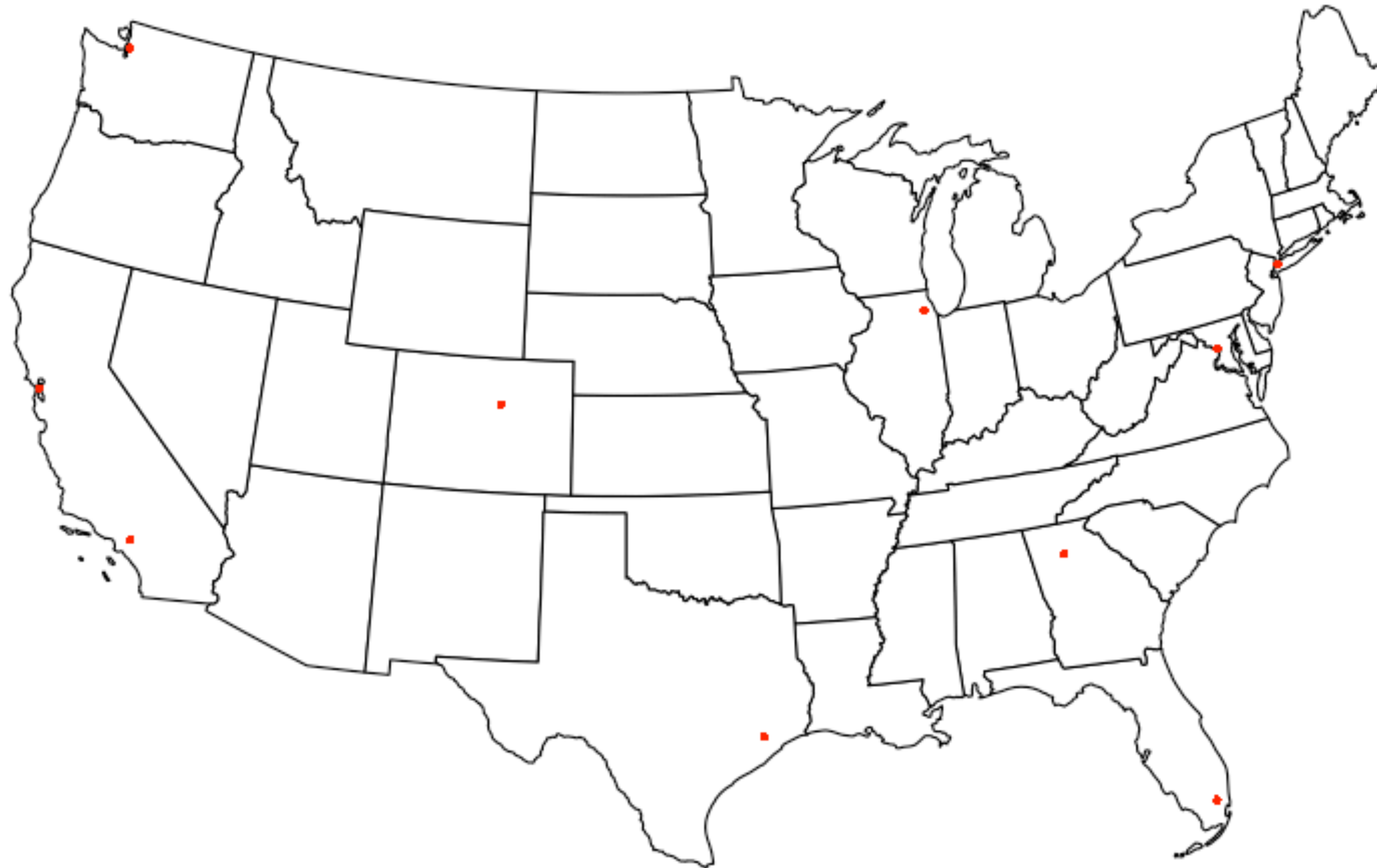
- **Tipos**

Escalamiento multidimensional clásico (lineal)

Escalamiento multidimensional métrico (no lineal)

Escalamiento multidimensional no métrico (no lineal)

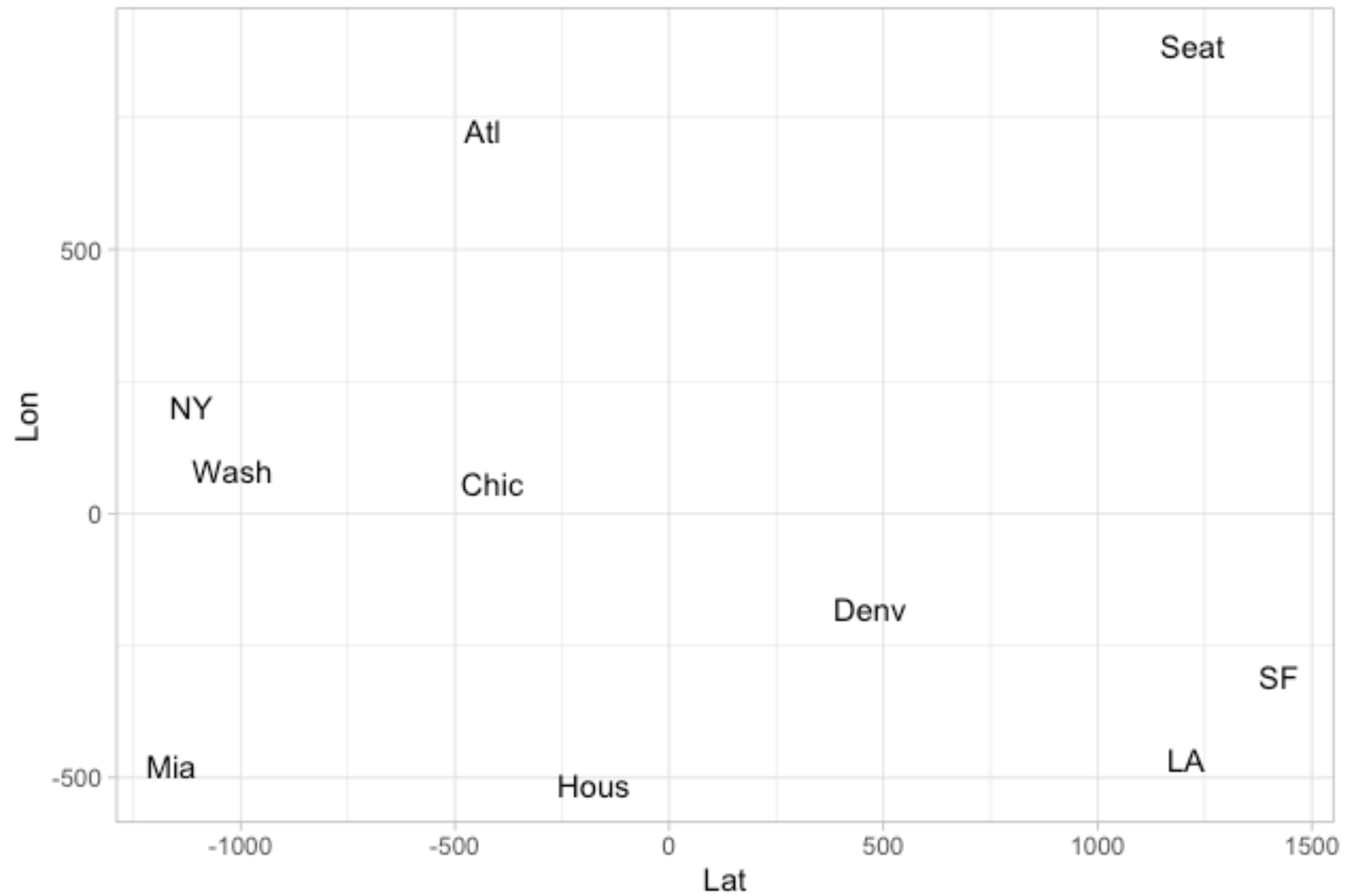
- Reconstrucción de un mapa a través de las distancias entre ciudades



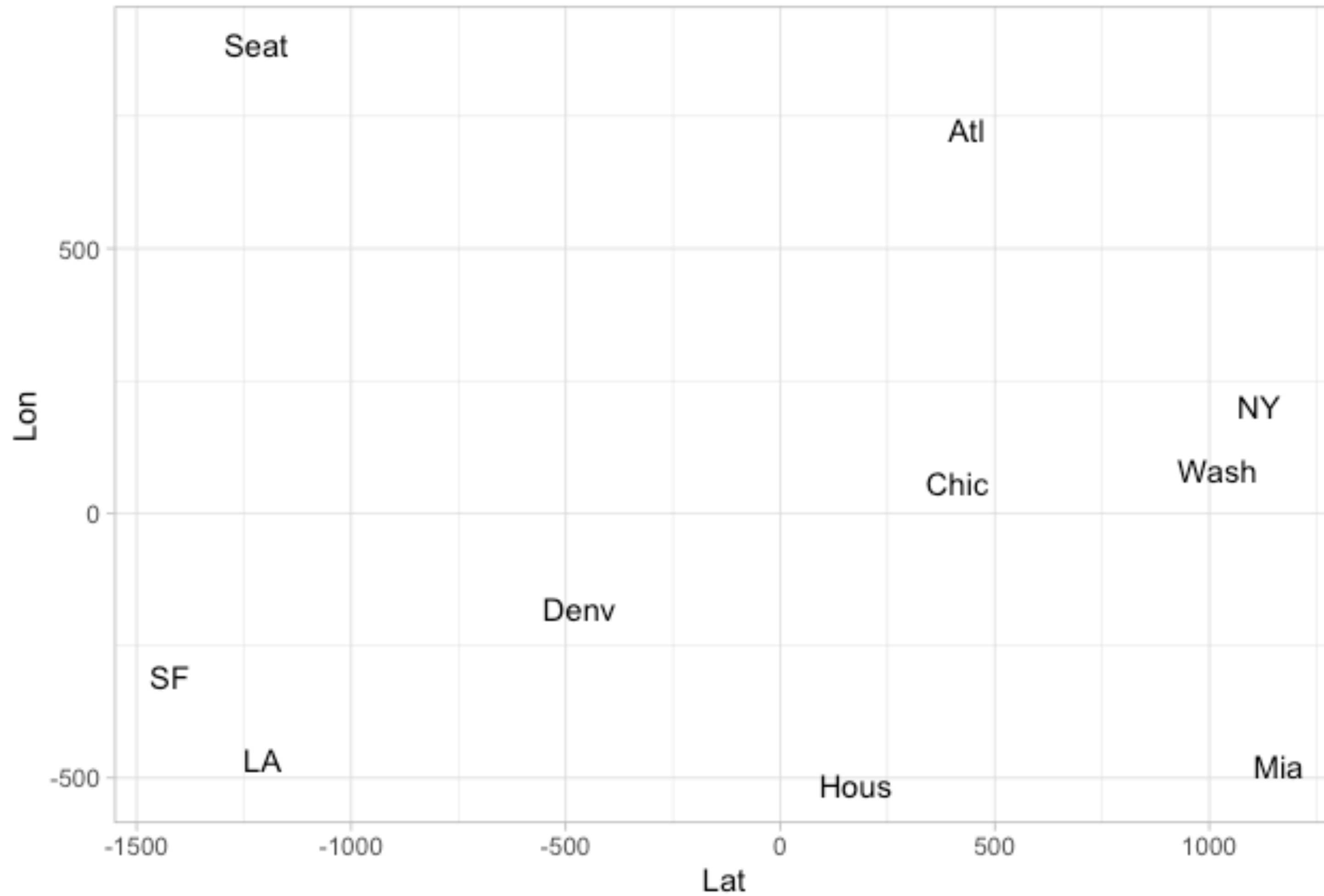
▸ Distancia en avión

	Atla	Chic	Denv	Hous	LA	Mia	NY	SF	Seat	Wash
Atlanta	-									
Chicago	587	-								
Denver	1212	920	-							
Houston	701	940	879	-						
LA	1936	1745	831	1374	-					
Miami	604	1188	1726	968	2339	-				
NY	748	713	1631	1420	2451	1092	-			
SF	2139	1858	949	1645	347	2594	2571	-		
Seattle	218	1737	1021	1891	959	2734	2408	678	-	
Wash. DC	543	597	1494	1220	2300	923	205	2442	2329	-

- Solución con escalamiento multidimensional clásico



- Rotamos la solución

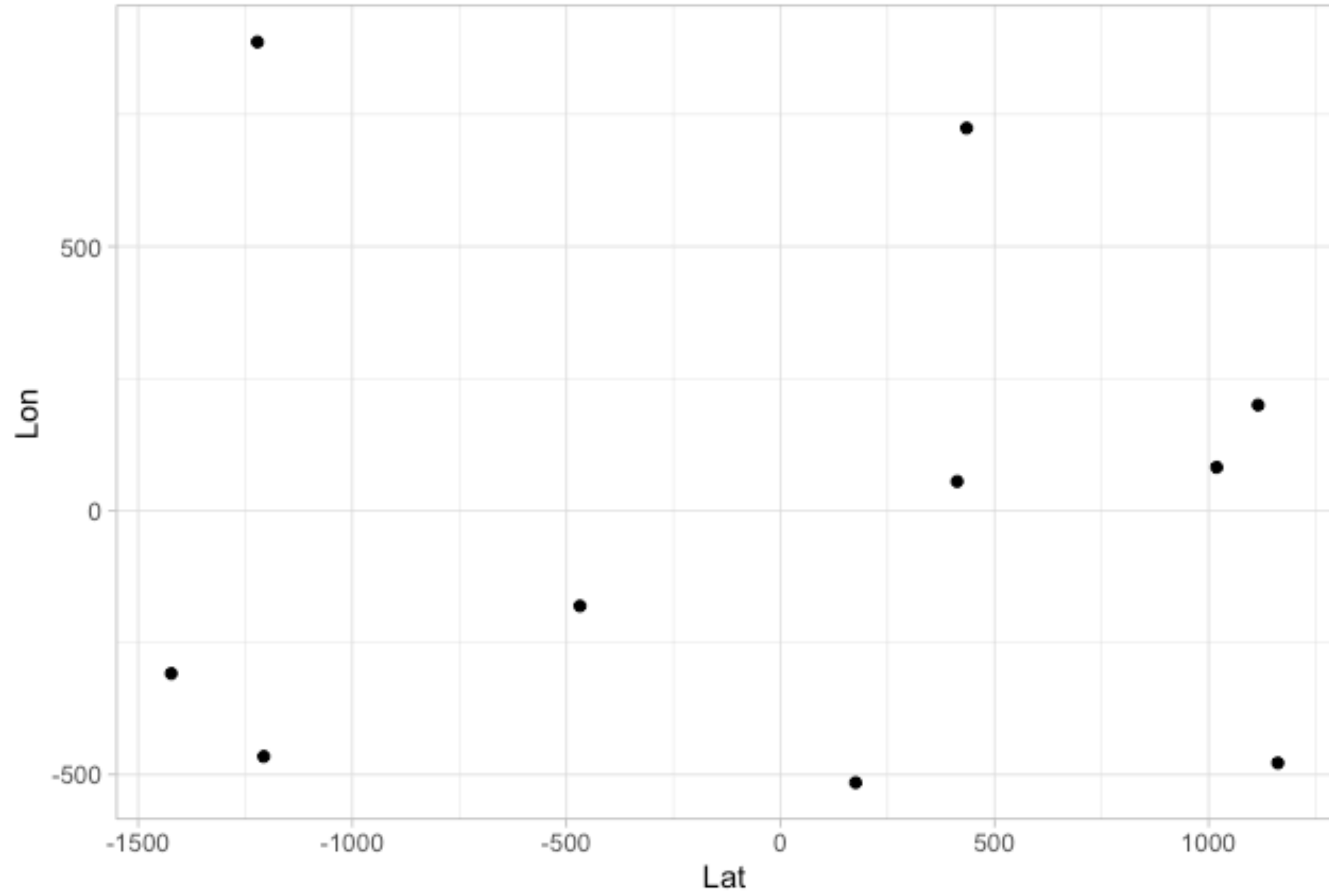


▸ Problema similar: identificar las ciudades

-										
587	-									
1212	920	-								
701	940	879	-							
1936	1745	831	1374	-						
604	1188	1726	968	2339	-					
748	713	1631	1420	2451	1092	-				
2139	1858	949	1645	347	2594	2571	-			
218	1737	1021	1891	959	2734	2408	678	-		
543	597	1494	1220	2300	923	205	2442	2329	-	



# Ejemplo 1



# Escalamiento Multidimensional Métrico

▸ Construir una matriz de distancias **D**

1.  $d_{ij} \geq 0 \quad \forall i, j = 1, \dots, n$

2.  $d_{ii} = 0$

3.  $\mathbf{D} = \mathbf{D}^T$

- Construir una matriz de distancias **D**
  1.  $d_{ij} \geq 0 \quad \forall i, j = 1, \dots, n$
  2.  $d_{ii} = 0$
  3. **D** = **D**<sup>T</sup>
- Encontrar un conjunto de vectores k-dimensionales tales que  $d(\mathbf{x}_i, \mathbf{x}_j) \approx \delta(\mathbf{y}_i, \mathbf{y}_j)$

- Construir una matriz de distancias **D**

1.  $d_{ij} \geq 0 \quad \forall i, j = 1, \dots, n$

2.  $d_{ii} = 0$

3.  $\mathbf{D} = \mathbf{D}^T$

- Encontrar un conjunto de vectores k-dimensionales tales que  $d(\mathbf{x}_i, \mathbf{x}_j) \approx \delta(\mathbf{y}_i, \mathbf{y}_j)$

- **Observaciones**

1. **D** es Euclidiana si existe una configuración p-dimensional tal que  $d_{ij} = \delta_{ij}$  (no siempre ocurre) y tenemos la solución clásica

2. En ocasiones **D** es una medición con error.

- **Definición (matriz doblemente centrada)**

Sea **D** una matriz de distancias entonces la matriz doblemente centrada está definida como

$$\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$$

- Donde:

$$\mathbf{A} = -\frac{1}{2}\mathbf{D} \odot \mathbf{D} \qquad a_{ij} = -\frac{d_{ij}^2}{2}$$

## ► Teorema

Sea  $\mathbf{D}_{n \times n}$  una matriz de distancias con matriz doblemente centrada  $\mathbf{B} = -\frac{1}{2}\mathbf{H}(\mathbf{D} \odot \mathbf{D})\mathbf{H}$

entonces

1. Si  $\mathbf{D}_{n \times n}$  es Euclidiana entonces  $\mathbf{B} = (\mathbf{H}\mathbf{X})(\mathbf{H}\mathbf{X})^T$  y así  $\mathbf{B}$  es semi-defnida positiva.

## ► Teorema

Sea  $\mathbf{D}_{n \times n}$  una matriz de distancias con matriz doblemente centrada  $\mathbf{B} = -\frac{1}{2}\mathbf{H}(\mathbf{D} \odot \mathbf{D})\mathbf{H}$

entonces

1. Si  $\mathbf{D}_{n \times n}$  es Euclidiana entonces  $\mathbf{B} = (\mathbf{H}\mathbf{X})(\mathbf{H}\mathbf{X})^T$  y así  $\mathbf{B}$  es semi-definida positiva.
2. Si  $\mathbf{B}$  es semi-definida positiva con eigenvalores  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$  y descomposición espectral  $\mathbf{B} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$  entonces

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}$$

es una matriz de datos de dimensión  $n \times k$  con matriz Euclidiana de distancias  $\mathbf{D}$ .



► **Propiedades**

1. La solución no es única (invariante ante cambios de origen, rotaciones y reflexiones).

► **Propiedades**

1. La solución no es única (invariante ante cambios de origen, rotaciones y reflexiones).
2.  $\bar{\mathbf{y}} = \mathbf{0}$

► **Propiedades**

1. La solución no es única (invariante ante cambios de origen, rotaciones y reflexiones).
2.  $\bar{\mathbf{y}} = \mathbf{0}$
3. Robusto ante perturbaciones [e.g. Sibson (1978, 1979, 1981) y Mardia (1978)]

► **Propiedades**

1. La solución no es única (invariante ante cambios de origen, rotaciones y reflexiones).
2.  $\bar{\mathbf{y}} = \mathbf{0}$
3. Robusto ante perturbaciones [e.g. Sibson (1978, 1979, 1981) y Mardia (1978)]
4. Si  $\lambda_1$  y  $\lambda_2$  son mucho más grandes que los restantes eigenvalores y los elementos de  $\mathbf{y}^{(1)}$  y  $\mathbf{y}^{(2)}$  son razonablemente diferentes entonces si  $\sum_{k=1}^2 (y_{ik} - y_{jk})^2 \approx d_{ij}^2$  se tiene una buena representación en 2 dimensiones.

## ► Propiedades

1. La solución no es única (invariante ante cambios de origen, rotaciones y reflexiones).
2.  $\bar{\mathbf{y}} = \mathbf{0}$
3. Robusto ante perturbaciones [e.g. Sibson (1978, 1979, 1981) y Mardia (1978)]
4. Si  $\lambda_1$  y  $\lambda_2$  son mucho más grandes que los restantes eigenvalores y los elementos de  $\mathbf{y}^{(1)}$  y  $\mathbf{y}^{(2)}$  son razonablemente diferentes entonces si  $\sum_{k=1}^2 (y_{ik} - y_{jk})^2 \approx d_{ij}^2$  se tiene una buena representación en 2 dimensiones.
5. Si la matriz es no Euclidiana podemos hacer uso de los primeros  $k$  eigenvalores si son positivos y grandes y se tiene una configuración razonable con  $(\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)})$ .

► **Algoritmo (a partir de  $\mathbf{D}$ )**

1. Construir la matriz  $\mathbf{A}$ .
2. Obtener la matriz doblemente centrada  $\mathbf{B}$ .
3. Obtener los  $k$  eigenvalores positivos más grandes y los eigenvectores asociados.
4. Si  $k = 2$  o  $k = 3$  se tiene una configuración que se puede graficar.

► **Algoritmo (a partir de **D**)**

1. Construir la matriz **A**.
2. Obtener la matriz doblemente centrada **B**.
3. Obtener los  $k$  eigenvalores positivos más grandes y los eigenvectores asociados.
4. Si  $k = 2$  o  $k = 3$  se tiene una configuración que se puede graficar.

► **En R :** `cmdscale( d , k=2 , add=F)`

▸ Distancia en avión de 10 ciudades de Estados Unidos

	Atla	Chic	Denv	Hous	LA	Mia	NY	SF	Seat	Wash
Atlanta	-									
Chicago	587	-								
Denver	1212	920	-							
Houston	701	940	879	-						
LA	1936	1745	831	1374	-					
Miami	604	1188	1726	968	2339	-				
NY	748	713	1631	1420	2451	1092	-			
SF	2139	1858	949	1645	347	2594	2571	-		
Seattle	218	1737	1021	1891	959	2734	2408	678	-	
Wash. DC	543	597	1494	1220	2300	923	205	2442	2329	-



## Ejemplo Revisitado

- Los eigenvalores de **B** están dados por

$$\lambda_1 = 9213705$$

$$\lambda_2 = 2199924$$

$$\lambda_3 = 1082863$$

$$\lambda_4 = 3322.361$$

$$\lambda_5 = 385.8824$$

$$\lambda_6 = -6.251781e - 09$$

$$\lambda_7 = -93.23115$$

$$\lambda_8 = -2168.535$$

$$\lambda_9 = -9090.644$$

$$\lambda_{10} = -1722963$$

- Por lo que **B** no es Euclidiana

## Ejemplo Revisitado

- Nos quedamos con los 5 eigenvalores positivos y construimos **Y**

-434.76	724.22	440.93	0.19	-0.01
-412.61	55.04	-370.93	4.40	12.68
468.20	-180.66	-213.57	30.41	-9.59
-175.58	-515.22	362.84	9.49	-4.86
1206.68	-465.64	56.53	1.34	6.81
-1161.69	-477.98	479.60	-13.80	2.28
-1115.56	199.79	-429.67	-29.40	-7.14
1422.69	-308.66	-205.52	-26.06	-1.98
1221.54	887.20	170.45	-0.07	0
-1018.90	81.90	-290.65	23.51	1.82

# Otras Consideraciones

1. Si se tienen similitudes con las siguientes condiciones:

- $s_{ij} \leq s_{ii}$
- $s_{ij} = s_{ji}$

1. Si se tienen similitudes con las siguientes condiciones:

- $s_{ij} \leq s_{ii}$
- $s_{ij} = s_{ji}$

Podemos transformarlo a una matriz de disimilitudes

$$d_{ij} = (s_{ii} - 2s_{ij} + s_{jj})^{\frac{1}{2}}$$

1. Si se tienen similitudes con las siguientes condiciones:

- $s_{ij} \leq s_{ii}$
- $s_{ij} = s_{ji}$

Podemos transformarlo a una matriz de disimilitudes

$$d_{ij} = (s_{ii} - 2s_{ij} + s_{jj})^{\frac{1}{2}}$$

2. En la práctica es usual el modelo  $d(\mathbf{x}_i, \mathbf{x}_j) \approx \delta(\mathbf{y}_i, \mathbf{y}_j) + a$

1. Si se tienen similitudes con las siguientes condiciones:

- $s_{ij} \leq s_{ii}$
- $s_{ij} = s_{ji}$

Podemos transformarlo a una matriz de disimilitudes

$$d_{ij} = (s_{ii} - 2s_{ij} + s_{jj})^{\frac{1}{2}}$$

2. En la práctica es usual el modelo  $d(\mathbf{x}_i, \mathbf{x}_j) \approx \delta(\mathbf{y}_i, \mathbf{y}_j) + a$  (additive constant problem)
3. Relación cercana entre MDS y PCA.

## Ejemplo 2

- 88 calificaciones de 5 exámenes a libro abierto o cerrado.

Lineal (C)	Estadística (C)	Probabilidad(A)	Finanzas (A)	Cálculo (A)
97	92	77	72	96
83	88	90	75	96
95	83	81	71	96
75	82	73	75	83
83	73	75	75	78
73	71	82	69	88
71	77	75	70	83

- En R usamos `prcomp(...,center=T)` con  $\hat{\Sigma} = S$

## Ejemplo 2

- Los eigenvalores de **S** son:

$$\lambda_1 = 60000.28$$

$$\lambda_2 = 17478.45$$

$$\lambda_3 = 9006.942$$

$$\lambda_4 = 7511.62$$

$$\lambda_5 = 2805.543$$

- Iguales a los 5 eigenvalores de **B** distintos de cero



## Ejemplo 2

- Aplicando la transformación a los alumnos 1,2,3,4,86,87 y 88 se tiene para dos componentes

Alumno	PCA1	PCA2
1	-66.28	-6.48
2	-63.60	6.79
3	-62.86	-3.26
4	-44.51	5.65
86	44.35	7.86
87	62.54	7.58
88	65.93	2.66

## Ejemplo 2

- Las primeras coordenadas con MDS de los alumnos 1,2,3,4,86,87 y 88 son:

Alumno	MDS1	MDS2
1	-66.28	6.48
2	-63.60	-6.79
3	-62.86	3.26
4	-44.51	-5.65
86	44.35	-7.86
87	62.54	-7.58
88	65.93	-2.66

1. Si **D** es Euclideana entonces el MDS clásico (o análisis de coordenadas principales) da los mismos resultados que PCA.

1. Si  $\mathbf{D}$  es Euclideana entonces el MDS clásico (o análisis de coordenadas principales) da los mismos resultados que PCA.
2. MDS es más flexible ya que acepta a las observaciones  $\mathbf{X}$ , una matriz de distancias (o disimilitudes)  $\mathbf{D}$ .

1. Si  $\mathbf{D}$  es Euclideana entonces el MDS clásico (o análisis de coordenadas principales) da los mismos resultados que PCA.
2. MDS es más flexible ya que acepta a las observaciones  $\mathbf{X}$ , una matriz de distancias (o disimilitudes)  $\mathbf{D}$ .
3. MDS es computacionalmente más demandante.

# Escalamiento Multidimensional No Métrico