Escalamiento Multidimensional

José A. Perusquía Cortés

Análisis Multivariado, Semestre 2025-II

Motivación

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

- **Tipos**

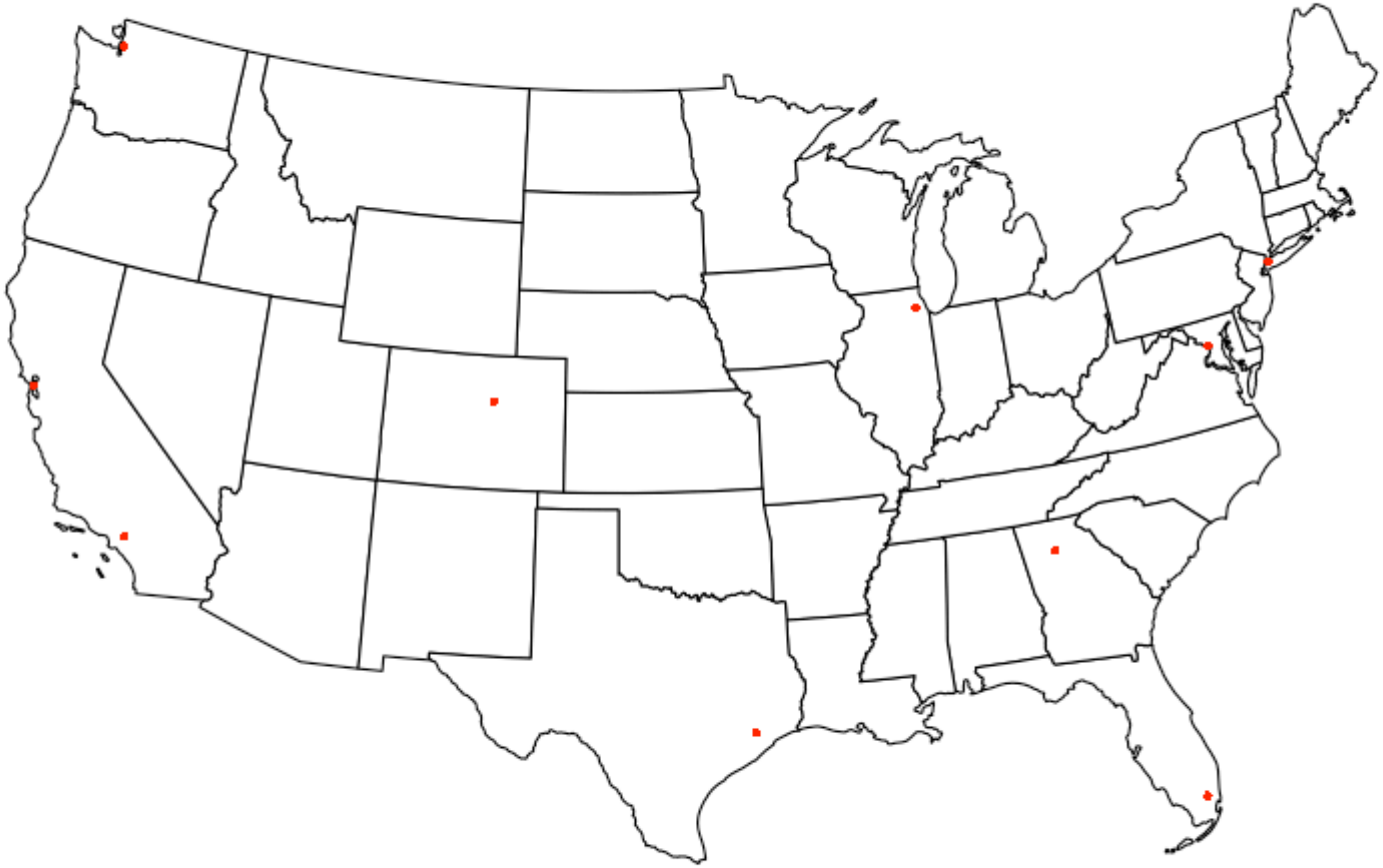
  Escalamiento multidimensional clásico (lineal)

  Escalamiento multidimensional métrico (no lineal)

  Escalamiento multidimensional no métrico (no lineal)

Ejemplo ciudades EE.UU.

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



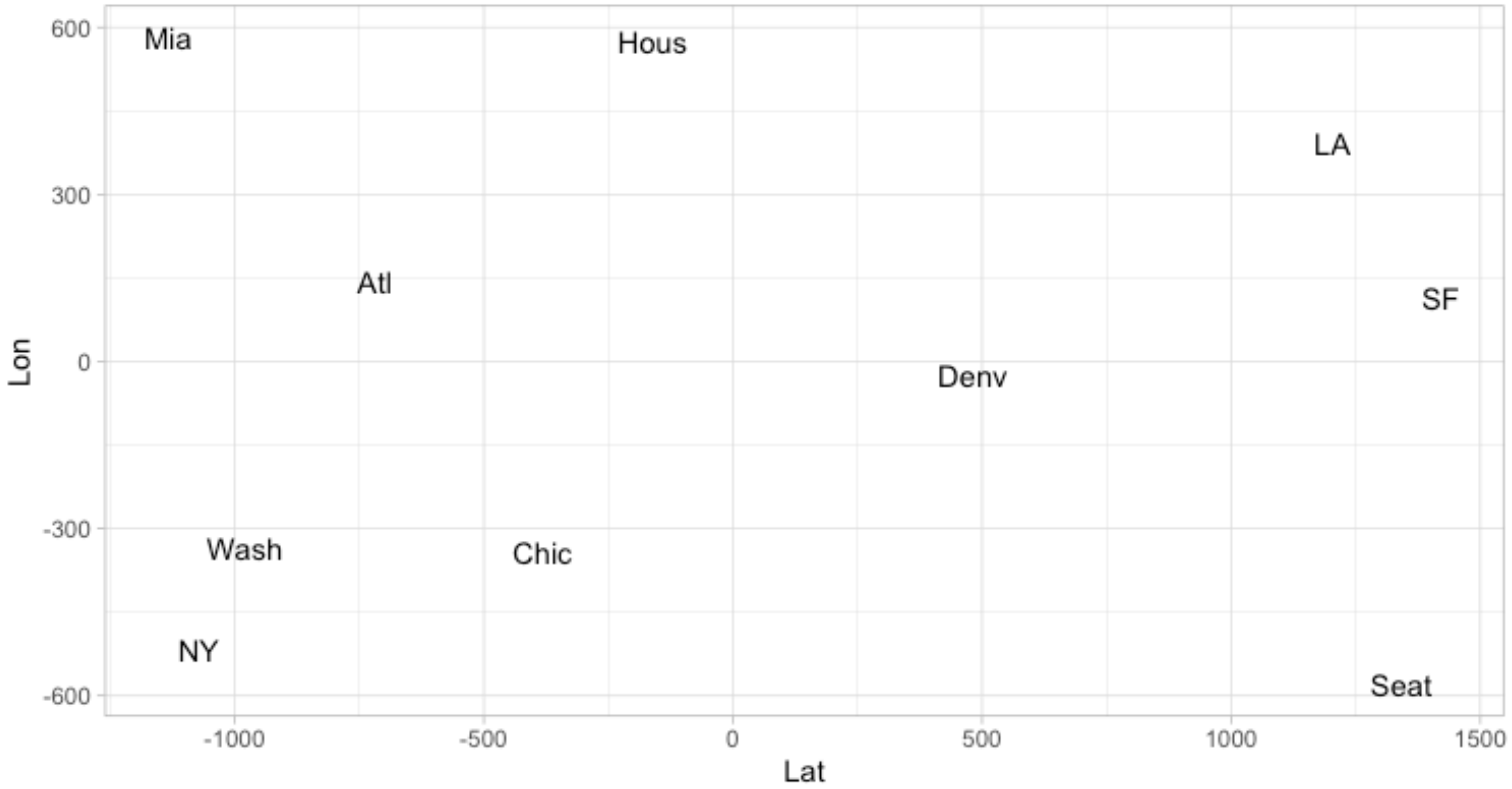
Ejemplo ciudades EE.UU.

- Distancia en avión

	Atl.	Chic.	Denv.	Houston	LA	Miami	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	2182	1737	1021	1891	959	2734	2408	678	-	
Wash.	543	597	1494	1220	2300	923	205	2442	2329	-

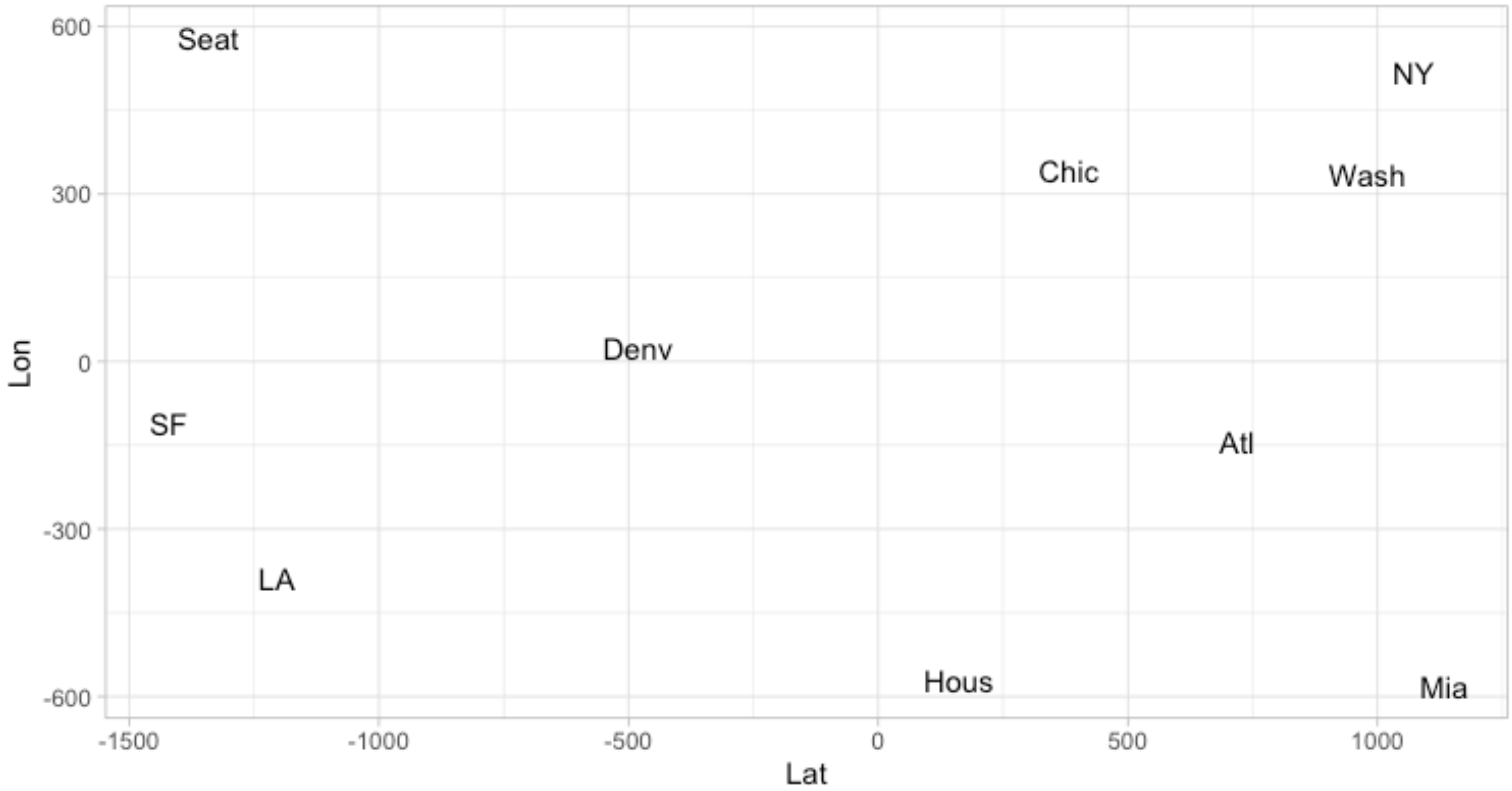
Ejemplo ciudades EE.UU.

- Utilizando el escalamiento multidimensional clásico



Ejemplo ciudades EE.UU.

- Rotando la solución

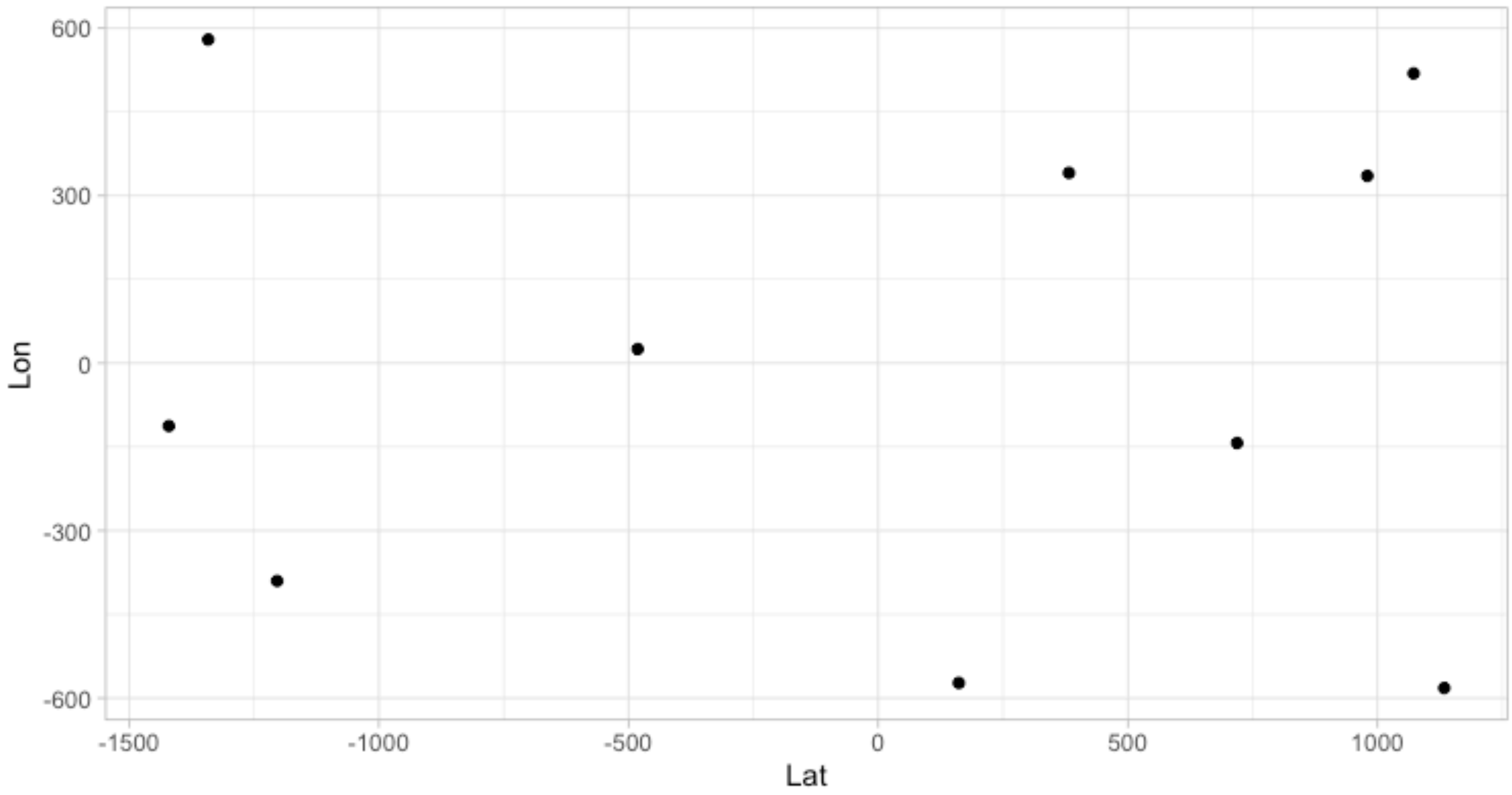


Ejemplo ciudades EE.UU.

- Problema similar: identificar las ciudades

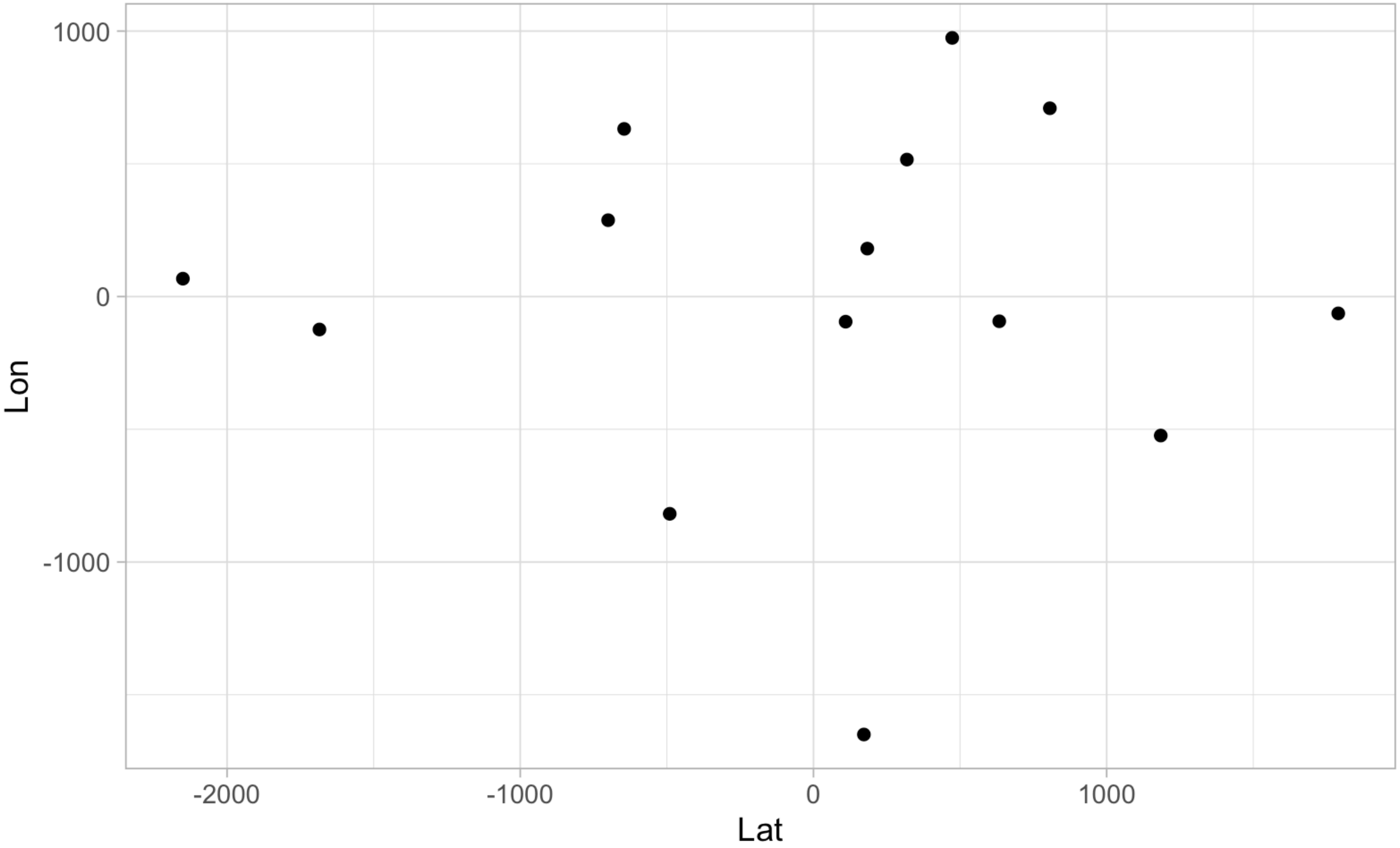
-									
587	-								
1212	920	-							
701	940	879	-						
1936	1745	831	1374	-					
604	1188	1726	968	2339	-				
748	713	1631	1420	2451	1092	-			
2139	1858	949	1645	347	2594	2571	-		
2182	1737	1021	1891	959	2734	2408	678	-	
543	597	1494	1220	2300	923	205	2442	2329	-

Ejemplo ciudades EE.UU.



Ejemplo capitales Europa

- ¡No siempre es fácil!



Escalamiento Multidimensional Métrico (Clásico)

Procedimiento

- Construir una matriz de distancias/disimilitudes $\mathbf{D}$

1. $d_{i,j} \geq 0$ para toda $i, j = 1, \ldots, n$
2. $d_{i,i} = 0$
3. $\mathbf{D} = \mathbf{D}^T$

- Encontrar un conjunto de vectores $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbb{R}^k$ tales que $d_{\mathbf{x}}(i, j) \approx d_{\mathbf{y}}(i, j)$

Observación 1

1. $\mathbf{D}$ es euclidiana si existe una configuración tal que $d_{\mathbf{x}}(i, j) = d_{\mathbf{y}}(i, j)$
2. En ocasiones $\mathbf{D}$ es una medición con error.

Construcción

Definición 1 (matriz doblemente centrada)

Sea $\mathbf{D}$ una matriz de “distancias” entonces la matriz doblemente centrada está definida como

$$\mathbf{B} = \mathbf{HAH}$$

donde,

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

Construcción

Teorema 1

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 $\mathbf{B} = (\mathbf{HX})(\mathbf{HX})^T$ y así $\mathbf{B}$ es semi-definida positiva.

2.Si $\mathbf{B}$ es semi-definida positiva con eigenvalores $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$ y descomposición espectral $\mathbf{B} = \mathbf{U\Lambda U}^T$ entonces

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

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)

2. $\bar{\mathbf{y}} = \mathbf{0}$

3. Robusto ante perturbaciones [e.g. Sibson (1978, 1979, 1981) y Mardia (1978)]

4. Si λ_1 y λ_2 son mucho más grandes que $\lambda_3, \ldots, \lambda_n$ y los elementos de $\mathbf{y}^{(1)}$ y $\mathbf{y}^{(2)}$ son diferentes entonces si $\sum_{k=1}^{2} (y_{ik} - y_{jk})^2 \approx d_{ij}^2$ se tiene una buena representación en $\mathbb{R}^2$

5. Si la matriz es no euclidiana podemos hacer uso de los primeros l eigenvalores positivos y así, se tiene una configuración razonable con $(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(l)})$

Algoritmo

1. Construir la matriz $\mathbf{A}$

2. Obtener la matriz doblemente centrada $\mathbf{B}$

3. Obtener los k valores propios positivos y los vectores propios asociados

4. Si $k = 2$ o $k = 3$ se tiene una configuración que se puede graficar

- **En R :** cmdscale( )

Ejemplo ciudades EE.UU.

- Distancia en avión de 10 ciudades de Estados Unidos

	Atl.	Chic.	Denv.	Houston	LA	Miami	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	2182	1737	1021	1891	959	2734	2408	678	-	
Wash.	543	597	1494	1220	2300	923	205	2442	2329	-

Ejemplo ciudades EE.UU.

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

$$\lambda_1 = 9582144$$
$$\lambda_2 = 1686820$$
$$\lambda_3 = 8157.298$$
$$\lambda_4 = 1432.87$$
$$\lambda_5 = 508.6687$$
$$\lambda_6 = 25.14349$$
$$\lambda_7 = -6.218108e-10$$
$$\lambda_8 = -897.7013$$
$$\lambda_9 = -5467.577$$
$$\lambda_{10} = -35478.89$$

- **D** no es Euclidiana

Ejemplo ciudades EE.UU.

- Nos quedamos con los 6 valores propios positivos y construimos $\mathbf{Y}$

$\mathbf{y}^{(1)}$	$\mathbf{y}^{(2)}$	$\mathbf{y}^{(3)}$	$\mathbf{y}^{(4)}$	$\mathbf{y}^{(5)}$	$\mathbf{y}^{(6)}$
-718.7594	142.99427	35.102499	-1.224963	-7.4094776	1.5046461
-382.0558	-340.83962	29.602228	-8.237885	-12.0242975	-2.3383016
481.6023	-25.28504	53.393802	1.339279	15.6658897	-0.9526963
-161.4663	572.76991	1.452571	-1.762318	-0.6718656	2.7007621
1203.7380	390.10029	-18.635065	14.974864	-3.1692006	-1.6561488
-1133.5271	581.90731	-32.268842	-2.375685	2.9718537	-2.0471878
-1072.2357	-519.02423	-34.341878	-14.253857	6.4473289	0.2709088
1420.6033	112.58920	-7.754755	-18.120276	-0.8054123	0.8695197
1341.7225	-579.73928	-23.650787	5.961453	-1.4286322	0.6143794
-979.6220	-335.47281	-2.899773	23.699388	0.4238136	1.0341183

- Podemos quedarnos con las primeras dos columnas

Otras consideraciones

1. Si se tienen similitudes con las siguientes condiciones:

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

   Se puede crear una matriz de disimilitudes

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

2.Relación cercana entre el escalamiento multidimensional clásico y los componentes principales

3.Se puede considerar la formulación: $d_{\mathbf{x}}(i,j) \approx d_{\mathbf{y}}(i,j) + a$ (additive constant problem)

MDS vs PCA

1. Si $\mathbf{D}$ es euclidiana 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 $\mathbf{X}$ o a una matriz de distancias/disimilitudes $\mathbf{D}$.

3. MDS es computacionalmente más demandante.

Ejemplo calificaciones

- Los eigenvalores de $\mathbf{S}$ son

$$
\begin{aligned}
\lambda_1 &= 60000.28 \\
\lambda_2 &= 17478.45 \\
\lambda_3 &= 9006.942 \\
\lambda_4 &= 7511.62 \\
\lambda_5 &= 2805.543
\end{aligned}
$$

- Iguales a los 5 eigenvalores de $\mathbf{B}$ distintos de cero

Ejemplo calificaciones

- Aplicando las transformaciones a los alumnos 1, 2, 3, 4, 86, 87 y 88

Alumno	PCA1	PCA2	MDS1	MDS2
1	-66.28	-6.48	-66.28	6.48
2	-63.60	6.79	-63.60	-6.79
3	-62.86	-3.26	-62.86	3.26
4	-44.51	5.65	-44.51	-5.65
86	44.35	7.86	44.35	-7.86
87	62.54	7.58	62.54	-7.58
88	65.93	2.66	65.93	-2.66

Constante aditiva

- Se busca encontrar la constante c más pequeña tal que un conjunto de disimilitudes tengan una representación euclidiana mediante,

$$d^c_{i,j} = d_{i,j} + c \qquad\qquad i \neq j$$

- Problema estudiado por muchos autores (e.g. Messick & Abelson, 1956; Saito, 1978; Cailliez, 1983)

- La solución de Cailliez se utiliza en **R:** `cmdscale(...,add=T)`

- Por cuestiones numéricas no se puede garantizar que todos los valores propios sean no negativos

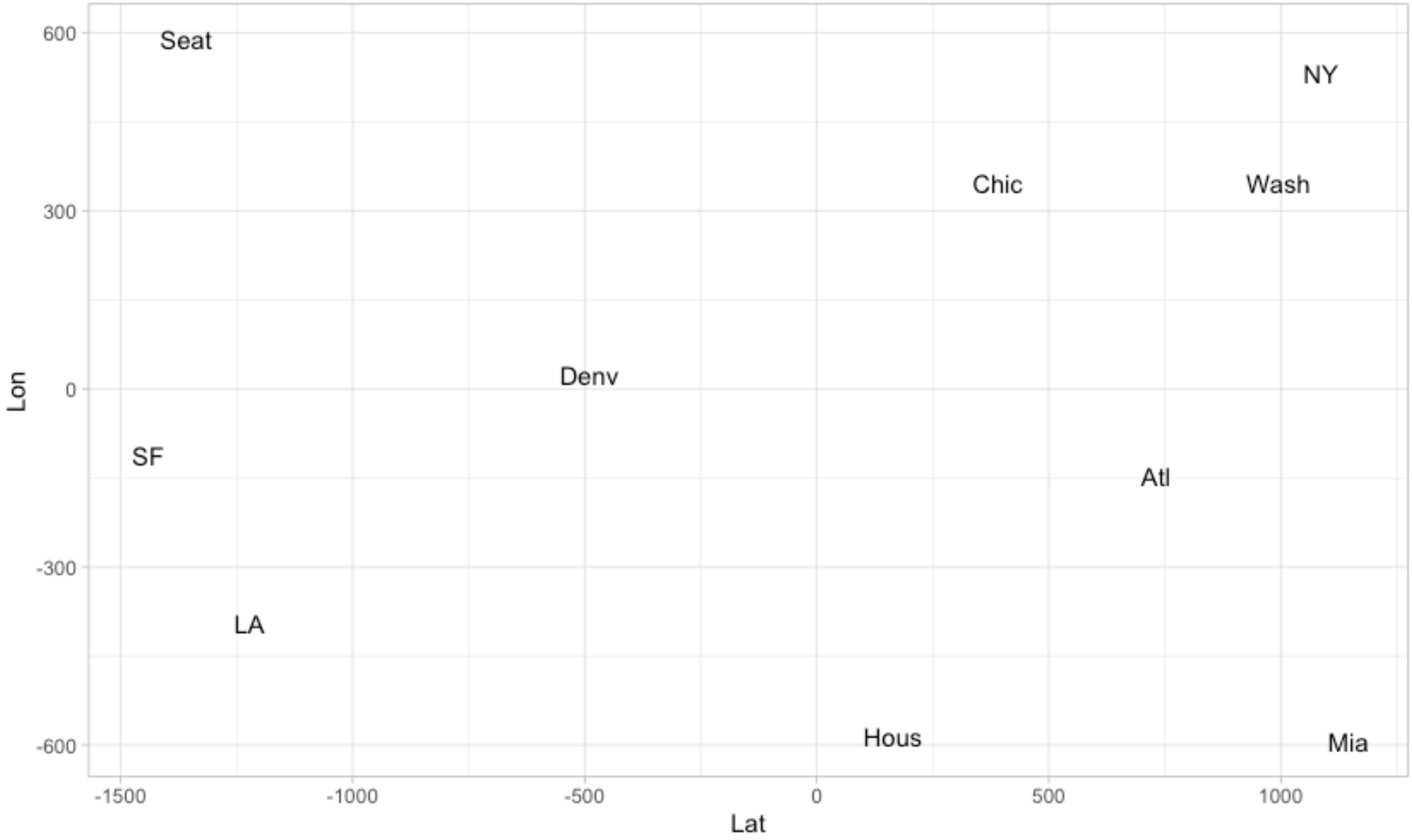
Ejemplo ciudades de EE.UU.

- Los valores propios de $\mathbf{B}$ al sumarle la constante aditiva $c = 39.12509$

$$
\begin{aligned}
\lambda_1 &= 9851759 \\
\lambda_2 &= 1760672 \\
\lambda_3 &= 49961.61 \\
\lambda_4 &= 23925.69 \\
\lambda_5 &= 22217.78 \\
\lambda_6 &= 15077.03 \\
\lambda_7 &= 11721.03 \\
\lambda_8 &= 7807.841 \\
\lambda_9 &= 1.55739e - 10 \\
\lambda_{10} &= -5.297162e - 10
\end{aligned}
$$

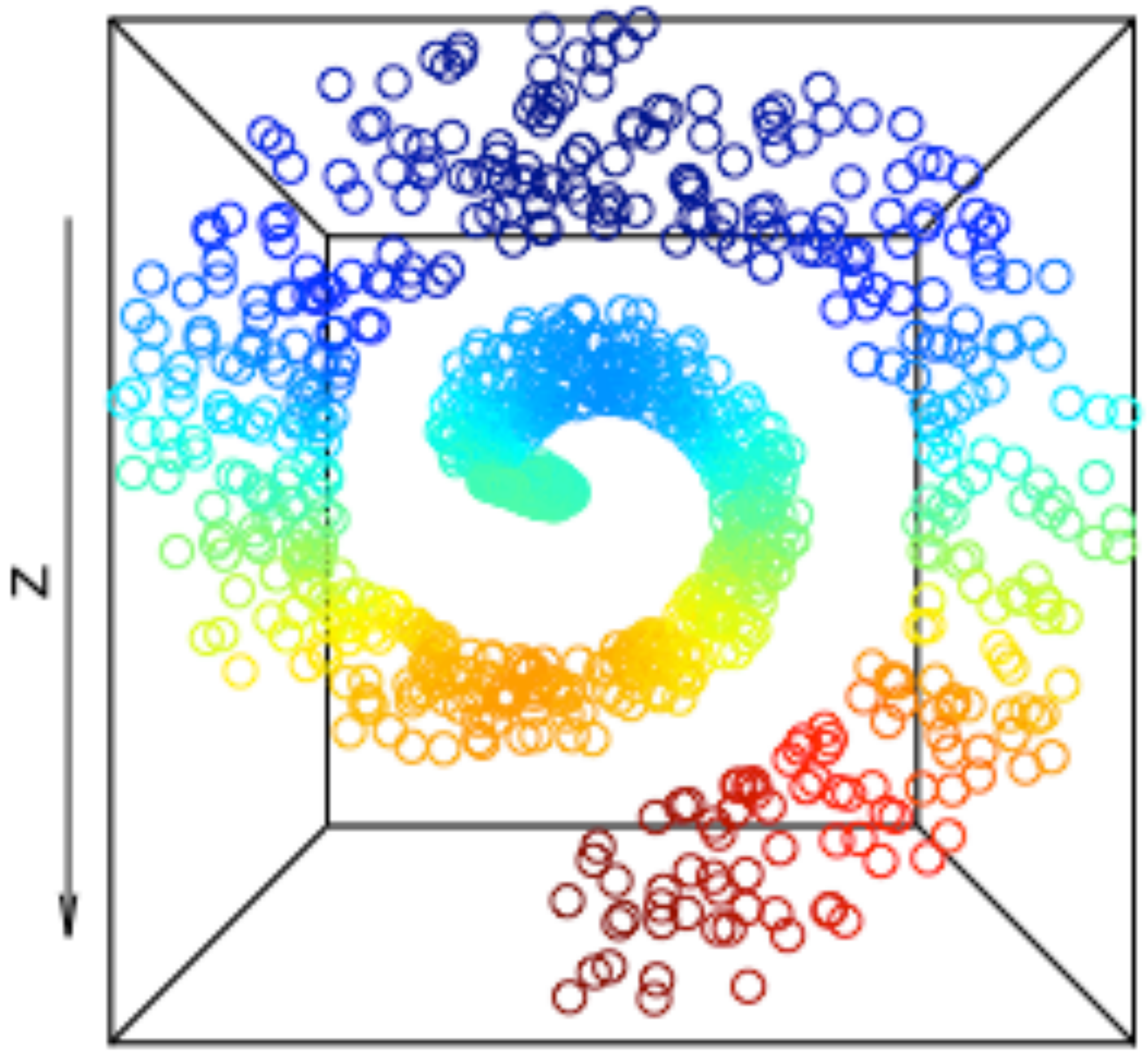
Ejemplo ciudades de EE.UU.

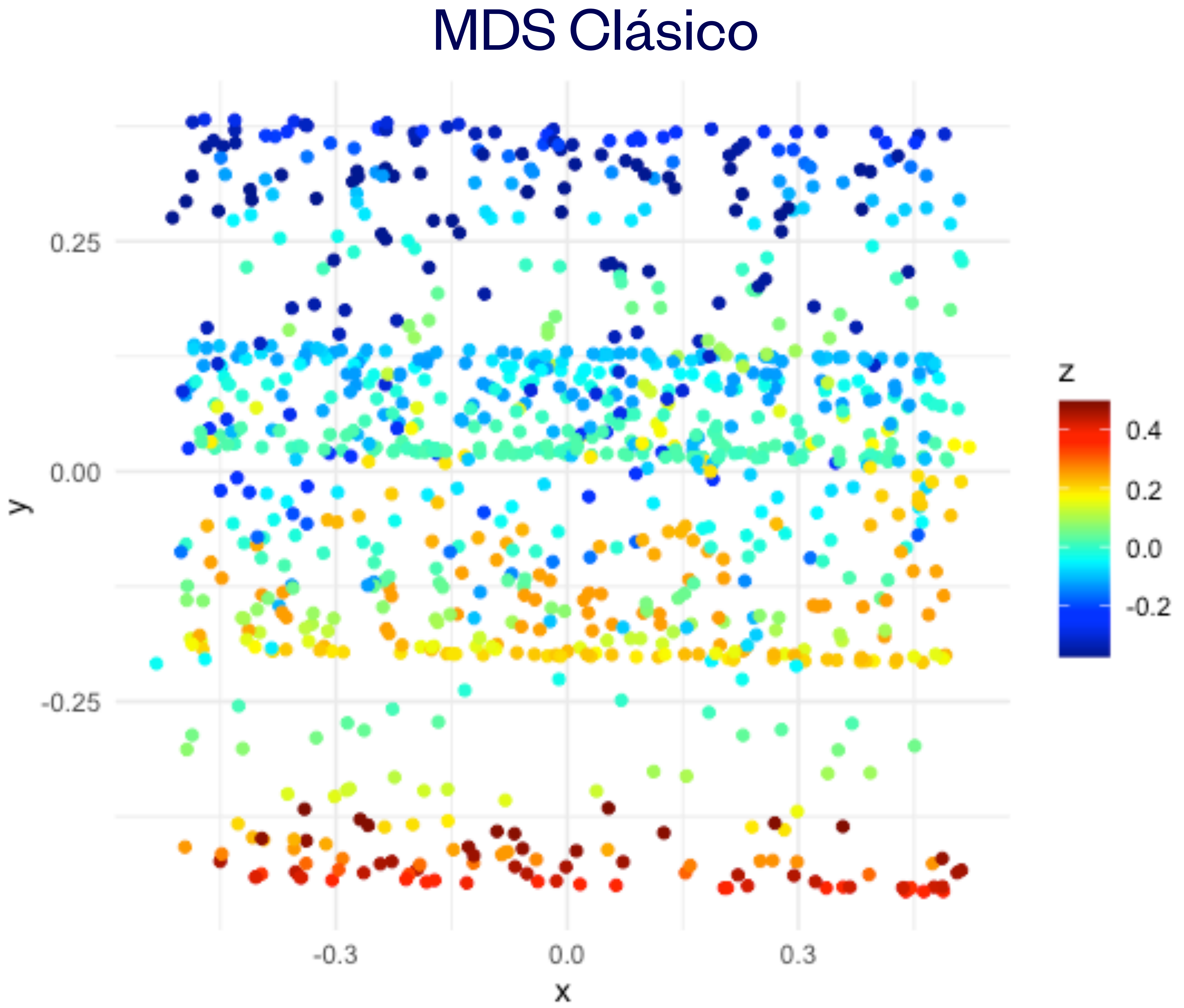
- El mapa reconstruido con la constante aditiva $c = 39.12509$



Escalamiento Multidimensional Métrico

Ejemplo rollo suizo (chocorol)





Motivación

- Una generalización no lineal del escalamiento clásico en donde se busca preservar las distancias y no solo los productos interiores

- **Objetivo**

  Minimizar una función objetivo conocida coloquialmente como “Stress”

$$\text{Stress} = \frac{1}{2} \sum_{i,j} w_{ij} \left[d_{\mathbf{x}}(i,j) - d_{\mathbf{y}}(i,j) \right]^2$$

- En la práctica, $w_{ij} = 1$ y $w_{ij} = 0$ (valores faltantes)

Sammon

- NLM: Mapeo no-lineal de Sammon (1969)

$$\text{Stress} = \frac{1}{c}\sum_{i,j}\frac{\left[d_{\mathbf{x}}(i,j) - d_{\mathbf{y}}(i,j)\right]^2}{d_{\mathbf{x}}(i,j)} \qquad c = \sum_{i<j} d_{\mathbf{x}}(i,j)$$

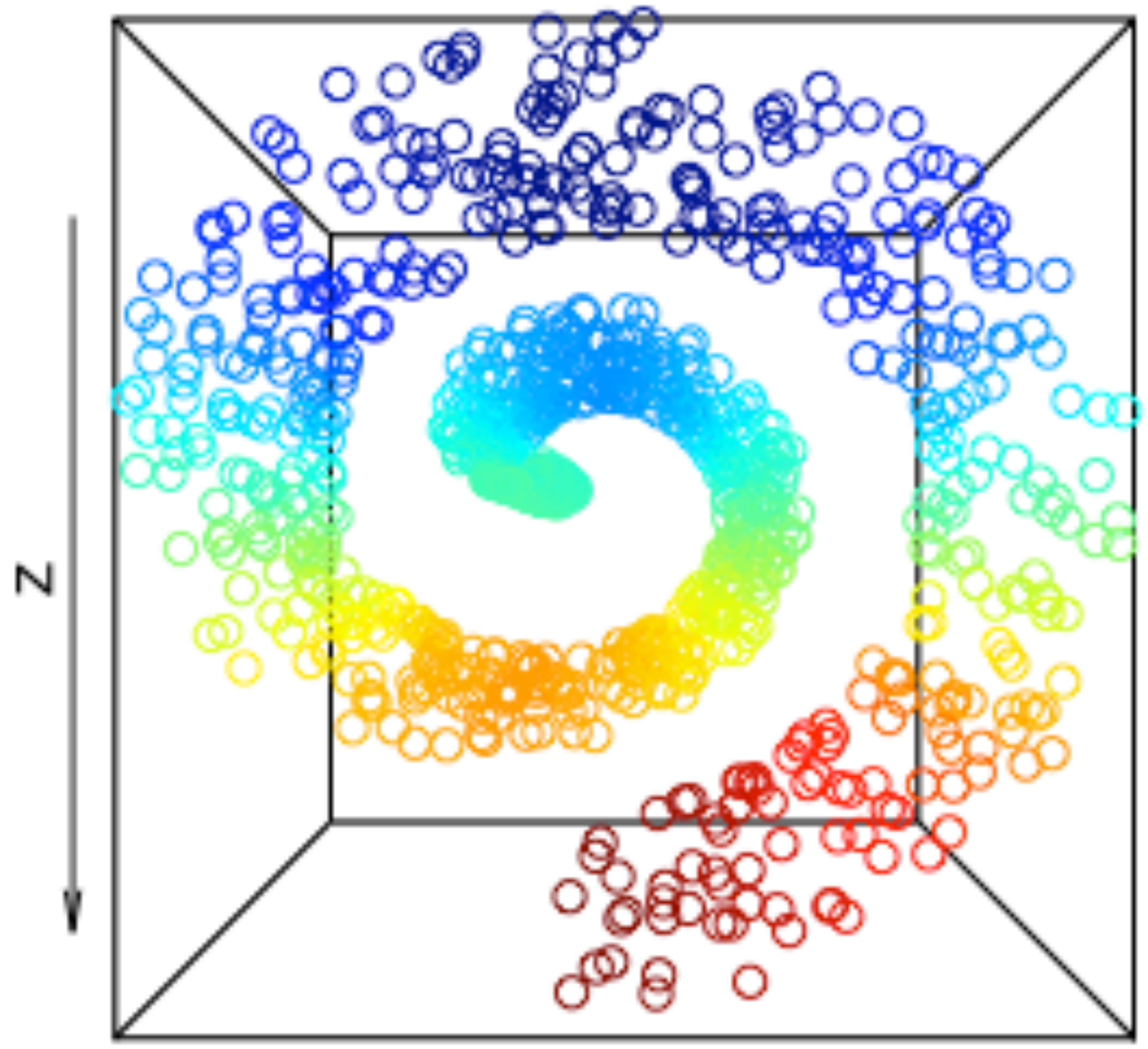
- Por lo general, $d_{\mathbf{x}}(i,j)$ es la distancia Euclidiana (no necesariamente)

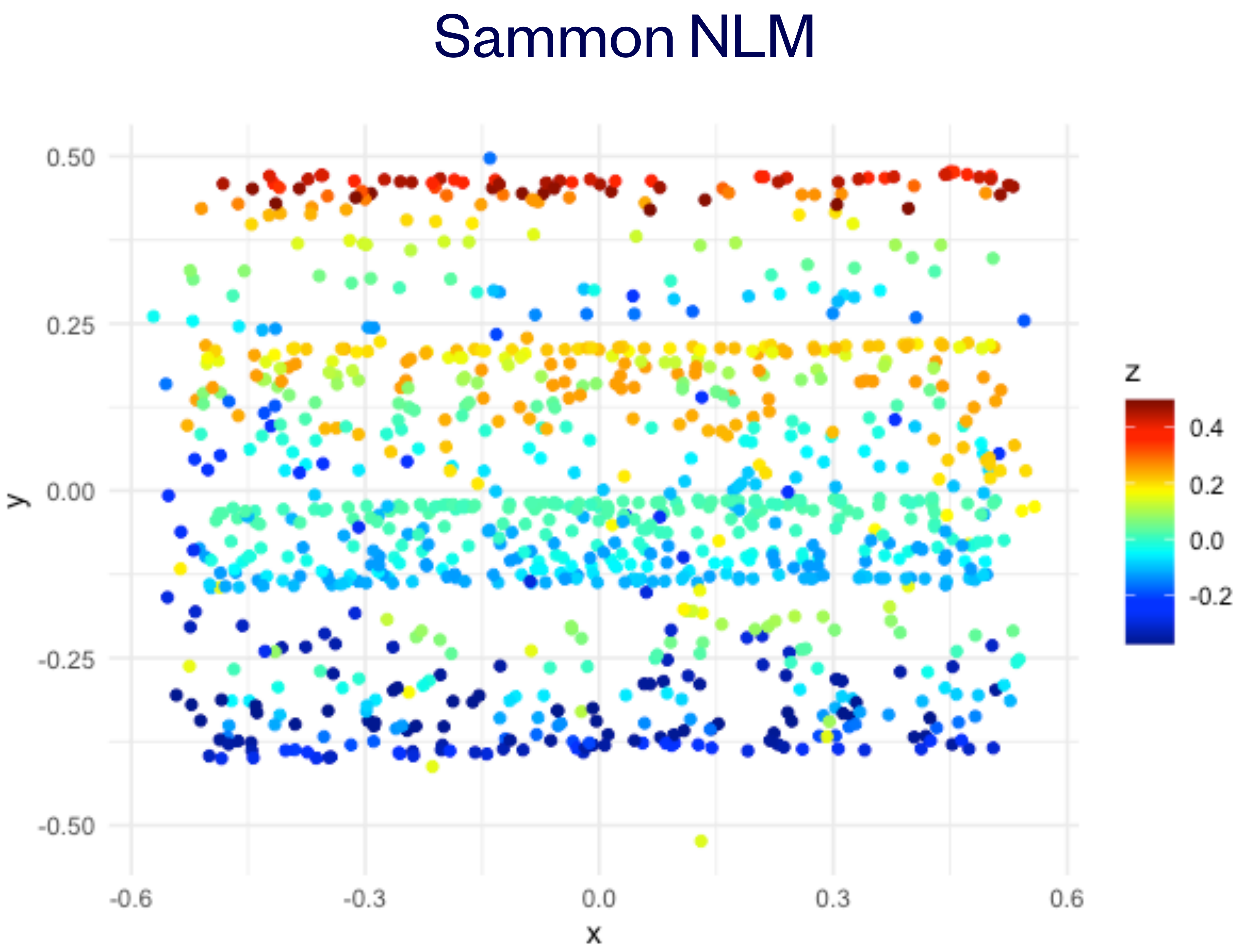
- Da más importancia a distancias cortas

- Requiere una rutina numérica (quasi-Newton) y de un parámetro "magic" (recomendado entre .3 y .4)

- En **R:** función sammon en librería MASS

Ejemplo rollo suizo (chocorol)





Escalamiento Multidimensional No Métrico

Motivación

- Alternativa menos rígida al MDS utilizando una función monótona desconocida de las distancias/proximidades, i.e.

$$d_x(i,j) = f(d_y(i,j))$$

- Para el MDS no métrico, construimos $d_{\mathbf{y}}(i,j)$ utilizando solo los rangos de $d_{\mathbf{x}}(i,j)$, e.g. para las ciudades de Estados Unidos usamos:

  - El viaje más corto es entre NY y Washington D.C.

  - El segundo viaje más corto es entre Seattle y Atlanta.

  ...

  - El viaje más largo es entre Seattle y Miami

MDS no métrico

- **Objetivo**

Optimizar la función “stress”

$$\text{Stress} = \sqrt{\frac{\sum_{ij} w_{ij} \left[f(\delta_{\mathbf{x}}(i,j)) - d_{\mathbf{y}}(i,j) \right]^2}{c}}$$

donde

- $\delta_{\mathbf{x}}(i,j)$ son proximidades
- f es una función monótona tal que $f(\delta_{\mathbf{x}}(i,j)) \approx d_{\mathbf{x}}(i,j)$ (distancia Euclidiana)
- c es un factor de escala
- w_{ij} son pesos no negativos como en el escalamiento multidimensional métrico

isoMDS

- Algoritmo dado por Shepard (1962) y Kruskal (1964)

1. Dada una matriz de disimilitudes $\mathbf{D}$ ordenar las entradas fuera de la diagonal.
2. Para una configuración k-dimensional, minimizar la función Stress dada por

$$\text{Stress} = \sqrt{\frac{\sum_{i<j} [d^*_{ij} - d_{\mathbf{y}}(i,j)]^2}{\sum_{i<j} d_{\mathbf{y}}(i,j)^2}}$$

con respecto a valores d^*_{ij} tal que d^*_{ij} esté relacionada de forma monótona con $d_{\mathbf{x}}(i,j)$,

i.e., $d_{\mathbf{x}}(i,j) < d_{\mathbf{x}}(k,l) \Rightarrow d^*_{ij} \leq d^*_{kl}$.

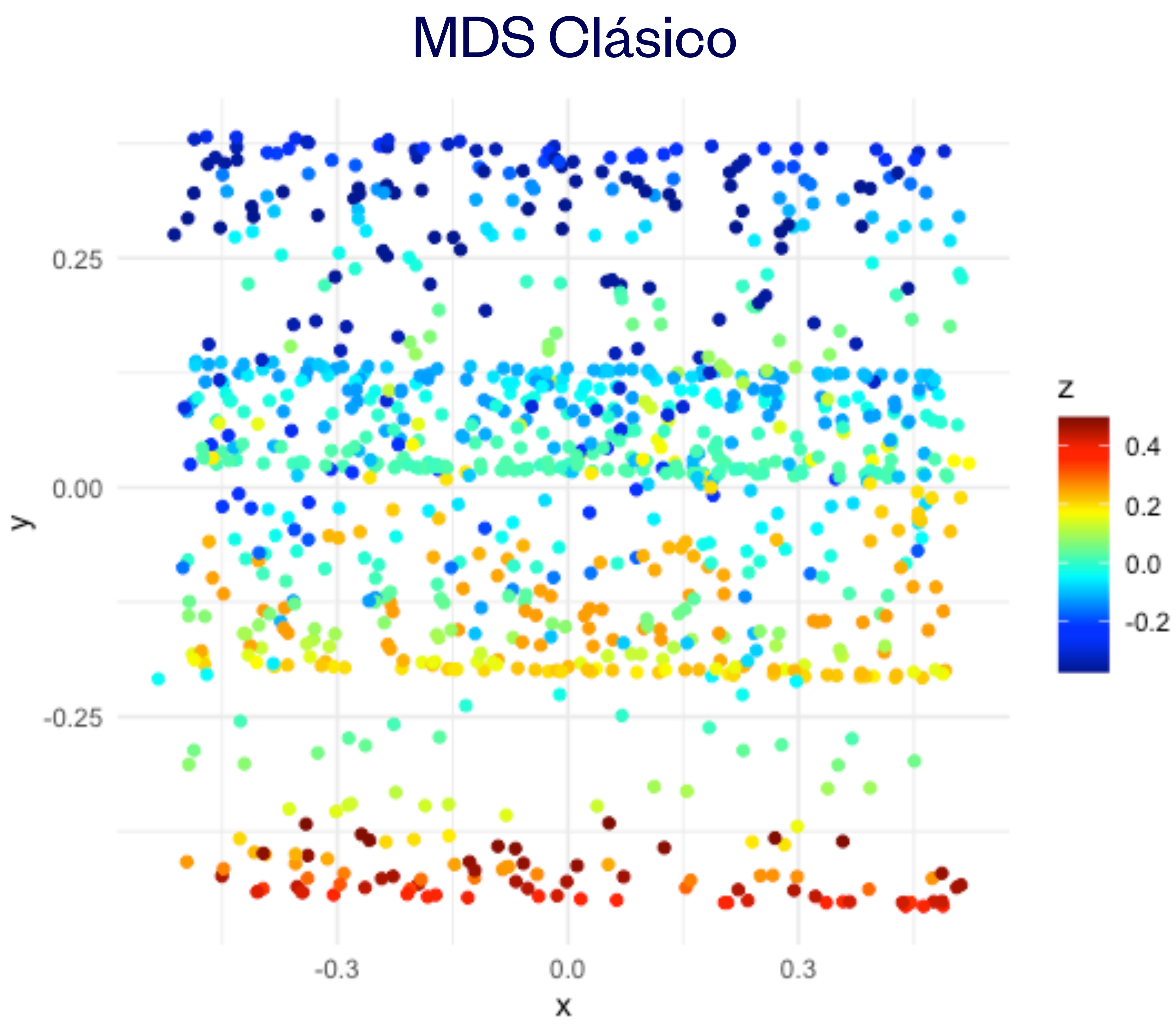
Consideraciones

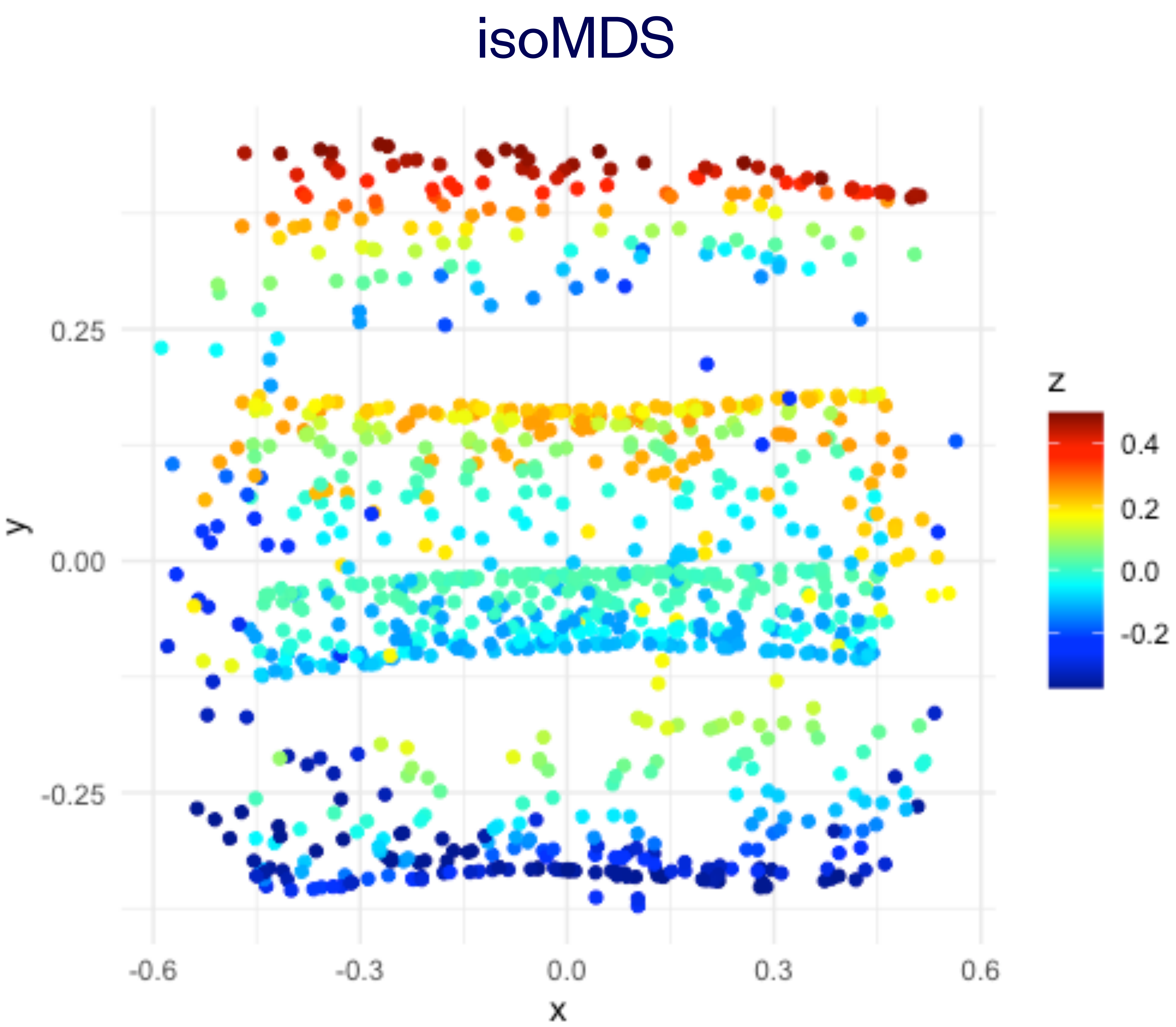
- Los d_{ij}^* se encuentran a través de una regresión monótona (isotonic regression).
- Requiere de rutinas numéricas.
- Para encontrar la dimensión adecuada calcular para cada k

$$S_k = \min \text{Stress}^2$$

detenerse hasta que S_k sea pequeño para $k = k_0$ o una regla de dedo de Kruskal donde $S_k \geq 20\,\%$ es pobre, $S_k = 10\,\%$ es justo, $S_k \leq 5\,\%$ es bueno y $S_k = 0$ es perfecto.
- En **R**: isoMDS/Shepard de la librería MASS utilizando una configuración inicial (e.g. solución clásica).

Ejemplo rollo suizo (chocorol)





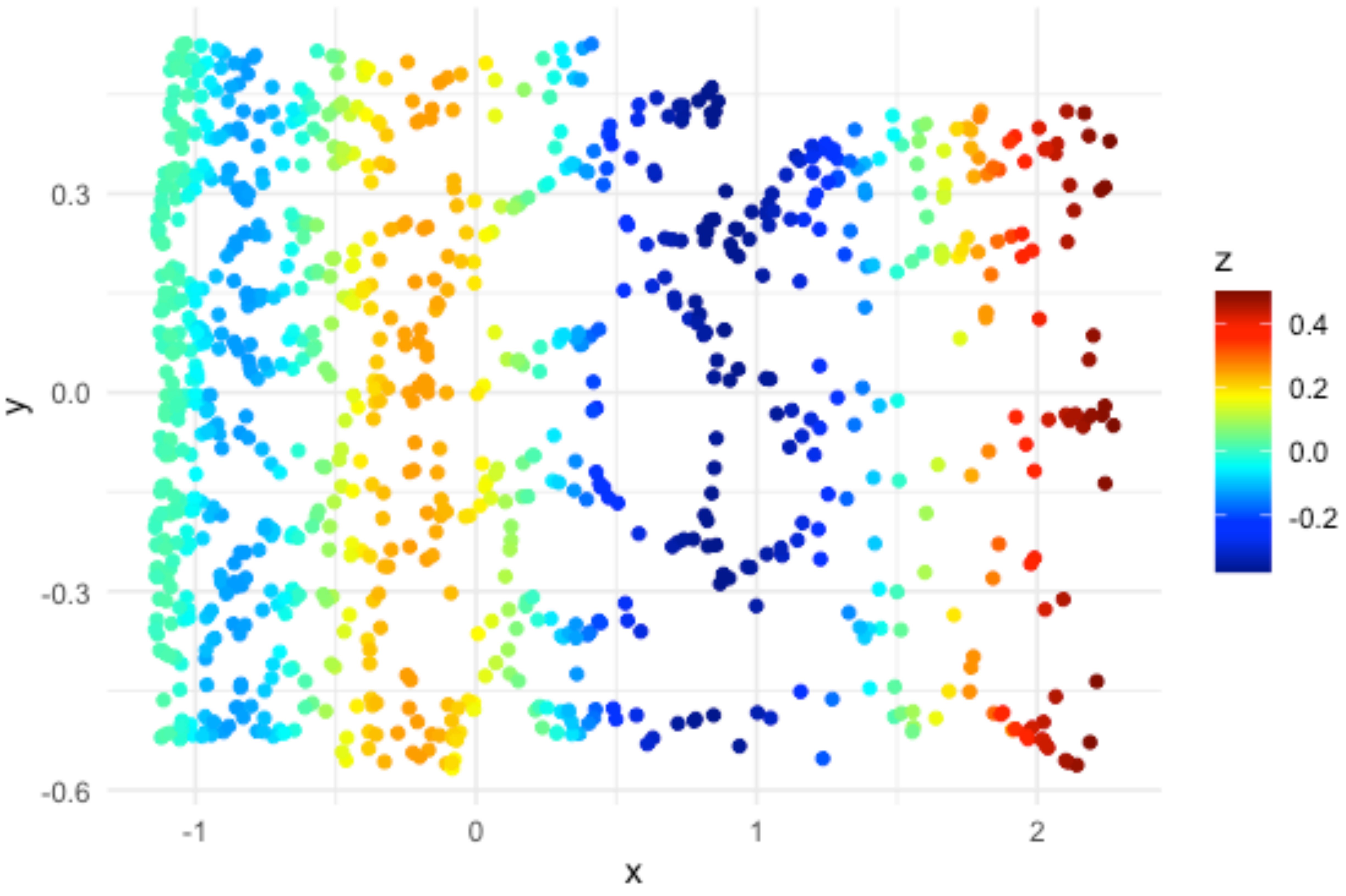
Variantes

- Hacer uso de otras distancias, e.g. distancia geodésica en la variedad
- Si la distancia geodésica es difícil de calcular (común) hacer uso de aproximaciones discretas usando grafos.
- Por ejemplo, Isomap (en **R** isomap en librería MASS):

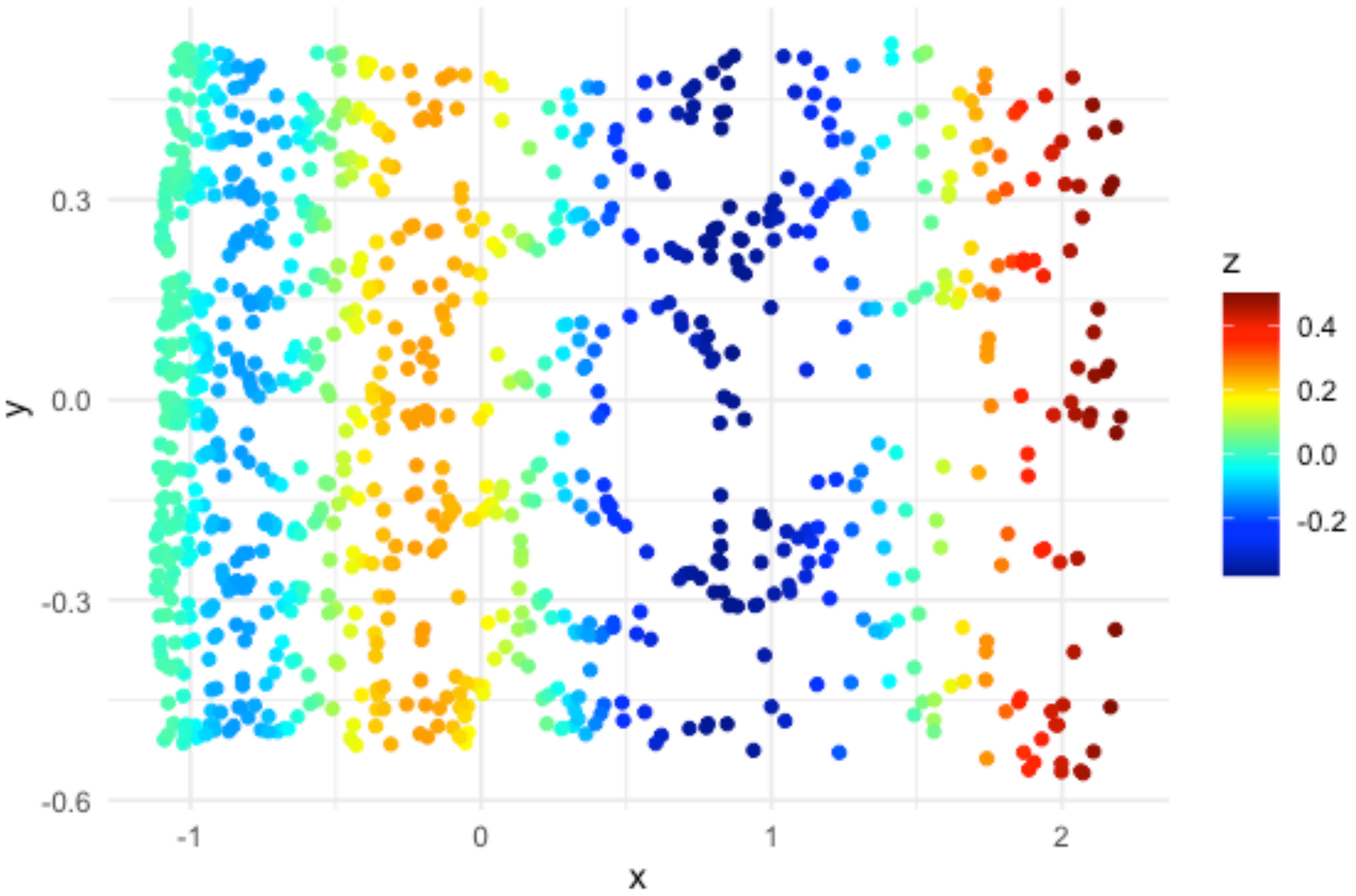
1. Conectamos cada punto con sus K vecinos más cercanos (o los que caigan en una bola de radio ϵ).
2. Aproximamos la matriz de distancias geodésicas a través del camino más corto en la red (algoritmo de Dijkstra o Floyd-Warshall)
3. Usamos escalamiento multidimensional clásico en la matriz de distancias.

Ejemplo rollo suizo (chocorol)

Isomap $K = 7$

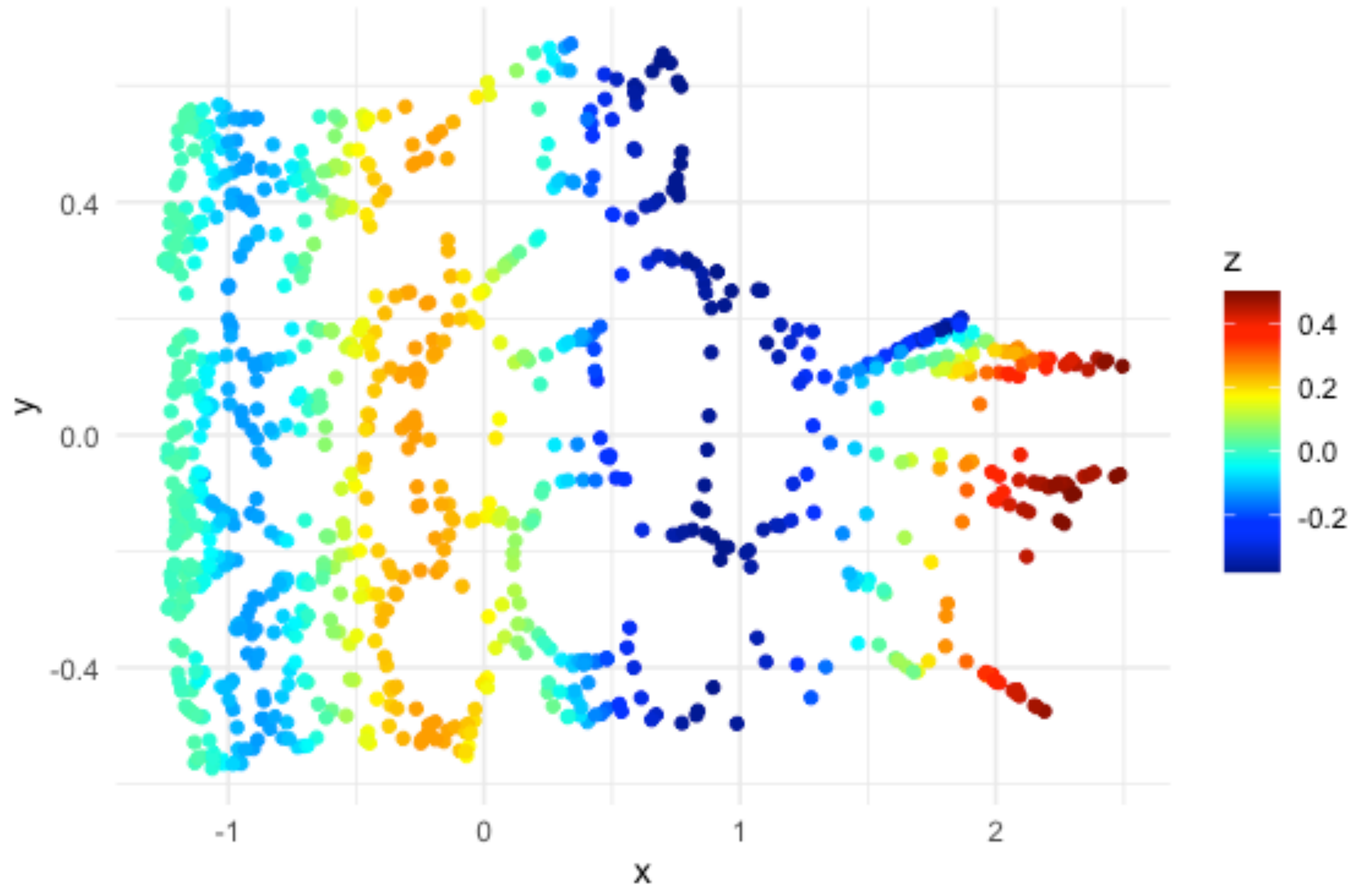


Isomap $K = 9$



Ejemplo rollo suizo (chocorol)

Isomap $K = 5$



Isomap $K = 10$