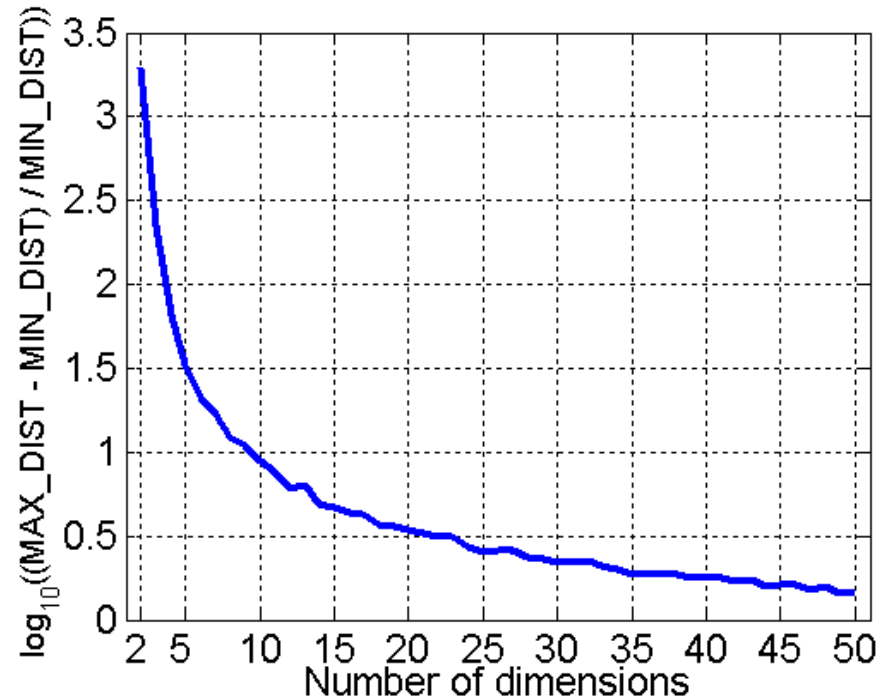


Büyük boyutun laneti (Curse of Dimensionality)

- Veri boyutu arttıkça örnekler (noktalar) uzay içinde çok fazla dağınık hale gelir.
- Noktaların yoğunluğu ya da aralarındaki uzaklık bir çok problem için çok önemlidir. Veri boyutu büyüdükçe yoğunluk ve uzaklık bilgisi anlamsızlaşır ve bu algoritmaların performansını etkiler.



- Rastgele 500 nokta üretelim
- Birbirine en uzak ve en yakın noktalar arası uzaklıkları ele alalım

Veri Küçültme

Boyut Küçültme

□ Amaç:

- Zaman ve hafıza gereksinimlerini azaltmak
- Kolay görselleştirme
- Alakasız öznitelikleri atmak ya da gürültü azaltmak

□ Yöntemler

- Temel Bileşen Analizi
 - Principle Component Analysis (PCA)
- Çok Boyutlu Ölçekleme
 - Multidimensional scaling (MDS)
- Diğerleri: eğitici yöntemler

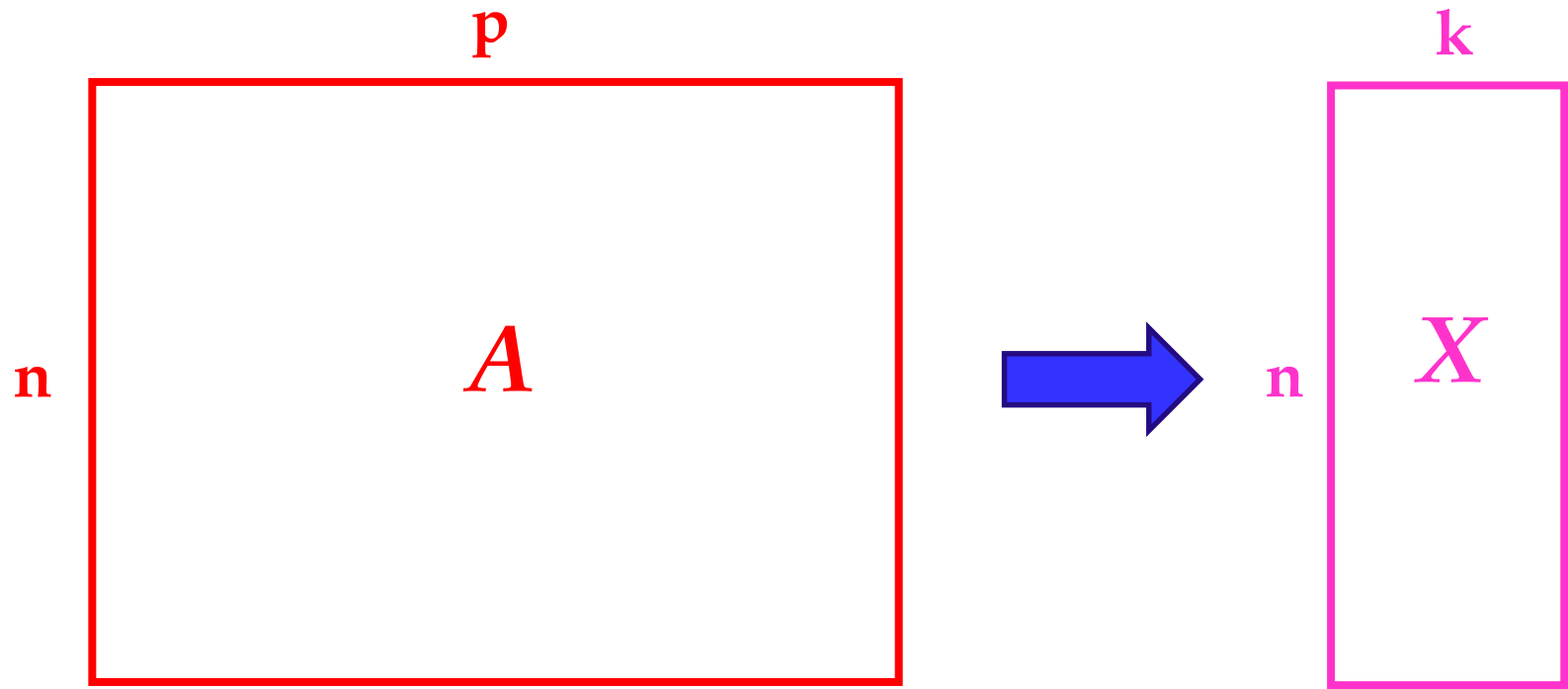
Boyut Küçültme

Temel Bileşen Analizi

- TBA tanımlayıcı bir yöntemdir
 - En eski yöntemlerden biridir
- Amaç
 - Büyük sayıda değişkenle ifade edilen örneklerin daha küçük uzayda temsili
 - Veri küçültme
 - Toplam varyansı en iyi açıklayan değişkenlerin tespiti
 - Yorumlama
- TBA sonuçları diğer algoritmalara girdi olabilir
 - regresyon
 - kümeleme
 - Sınıflandırma ve diğerleri

Boyut Küçültme

Temel Bileşen Analizi

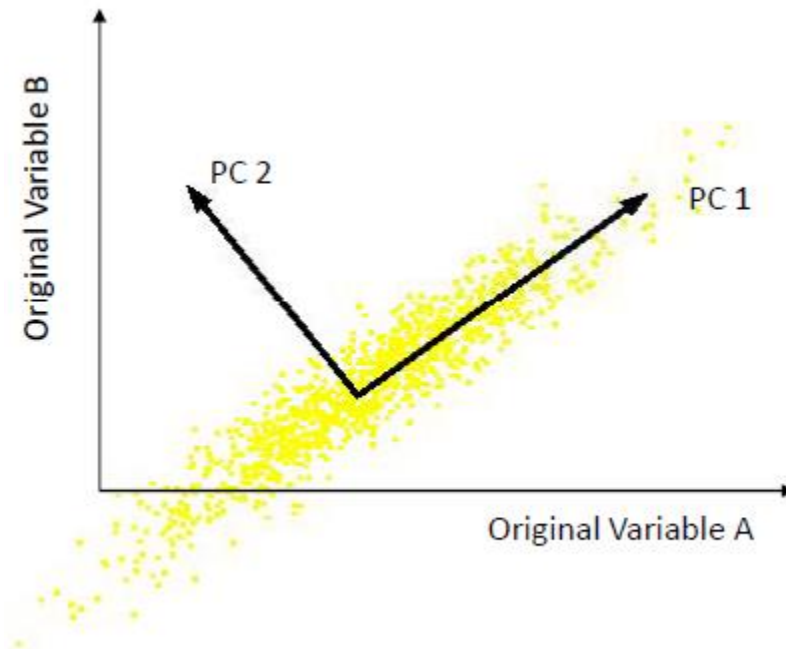


Öyle bir dönüşüm yapalım ki verideki varyansı en iyi şekilde saklayabilelim

Boyut Küçültme

Temel Bileşen Analizi

- Varyansın en yüksek olduğu birbirine dik eksenleri bulmak
 - PC1 yönü verinin daha çok değiştiği yönlerden biri

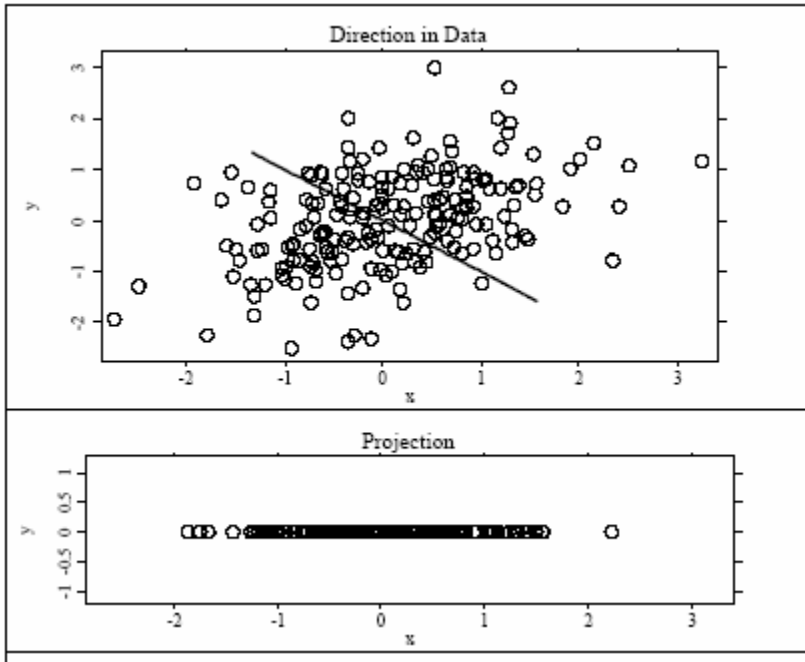


Boyut Küçültme

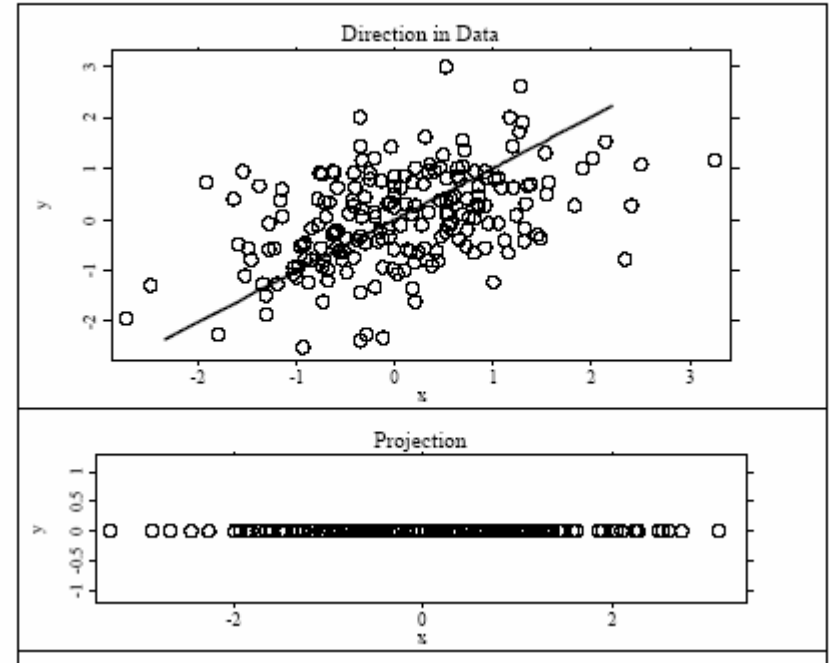
Temel Bileşen Analizi

□ Geometrik yorum

İyi



Daha İyi



Boyut Küçültme

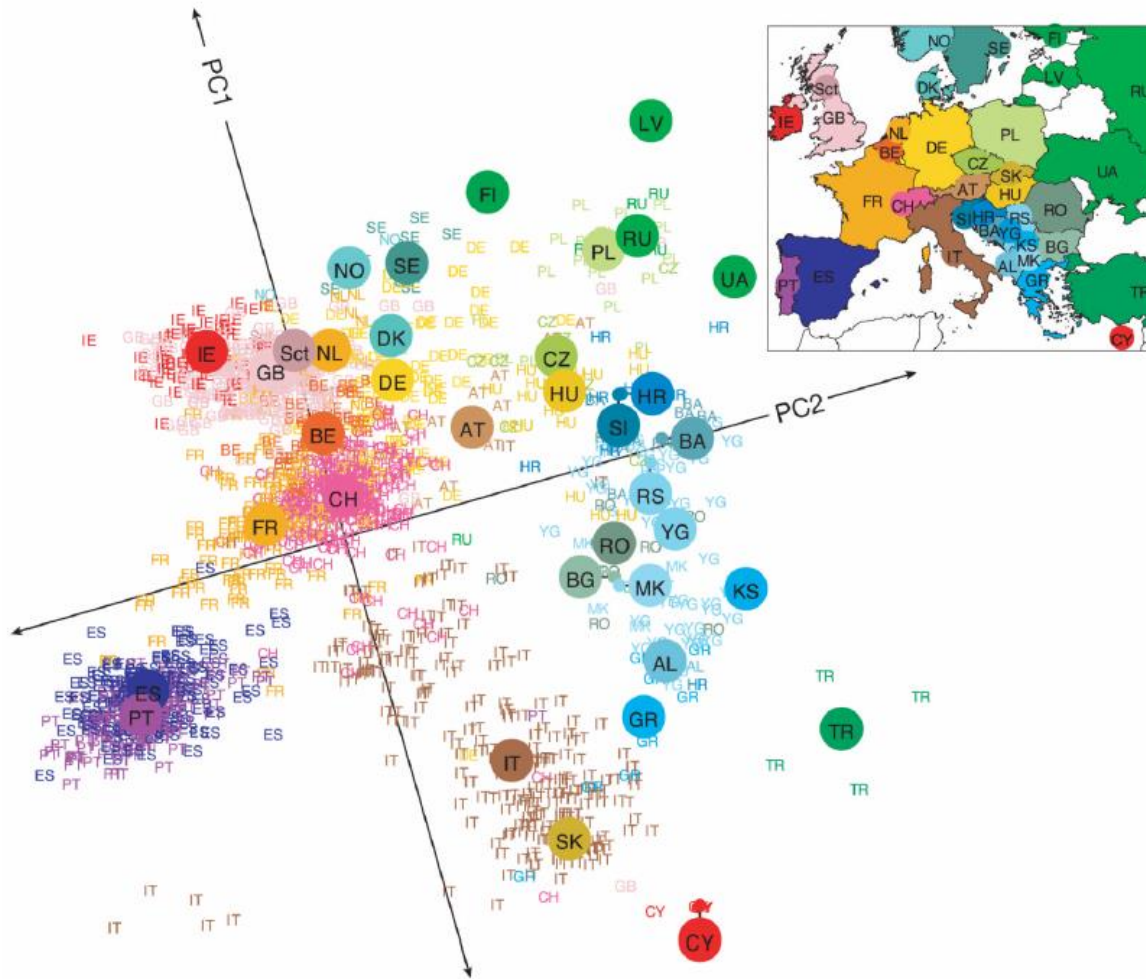
Temel Bileşen Analizi

▣ R kodları

Bir uygulama

“Genes mirror geography within Europe”

<http://www.nature.com/nature/journal/v456/n7218/full/nature07331.html>



The PC axes are rotated to emphasize the similarity to the geographic map of Europe. AL, Albania; AT, Austria; BA, Bosnia-Herzegovina; BE, Belgium; BG, Bulgaria; CH, Switzerland; CY, Cyprus; CZ, Czech Republic; DE, Germany; DK, Denmark; ES, Spain; FI, Finland; FR, France; GB, United Kingdom; GR, Greece; HR, Croatia; HU, Hungary; IE, Ireland; IT, Italy; KS, Kosovo; LV, Latvia; MK, Macedonia; NO, Norway; NL, Netherlands; PL, Poland; PT, Portugal; RO, Romania; RS, Serbia and

Boyut Küçültme

Temel Bileşen Analizi

□ Avantajları

- Çok basittir, parametresi neredeyse yoktur
 - Kaç bileşen tutulacağı dışında
- Veriyi aralarında korelasyon bulunmaya daha küçük sayıda öznitelik ile ifade eder
 - Veri sıkıştırma

□ Dezavantajları

- Numerik veri ve Normal (Gaus) dağılım
- Değişkenler arası ilişkilerin doğrusal (lineer) olduğunu varsayar
- Eğer ilişkiler doğrusal değilse
 - Bulunan temel eksenler anlamsızdır.
 - Kernel PCA denen yöntemler ile doğrusal olmayan TBA yapılabilir.

Boyut Küçültme

Çok Boyutlu Ölçekleme (MDS)

- MDS ve kümeleme analizi alakalıdır.
 - Genellikle parametrik olmayan, altında model barındırmayan, tanımlayıcı bir yöntemdir.
 - ~ doğrusal olmayan temel bileşen analizi de denebilir
- Veriyi daha küçük bir uzayda öyle bir şekilde ifade edelim ki asıl uzaydaki benzerlik bilgisi en iyi şekilde korunsun.
 - Çoğunlukla görselleştirme için kullanılır.
 - Tukey: "A picture is worth a thousand words"
 - *Given: an $n \times n$ matrix $\Delta = (\delta_{ij})$ of dissimilarities*
 - *Find: an $n \times k$ matrix \mathbf{X} of coordinates in a k -dimensional space, such that distances \approx dissimilarities*

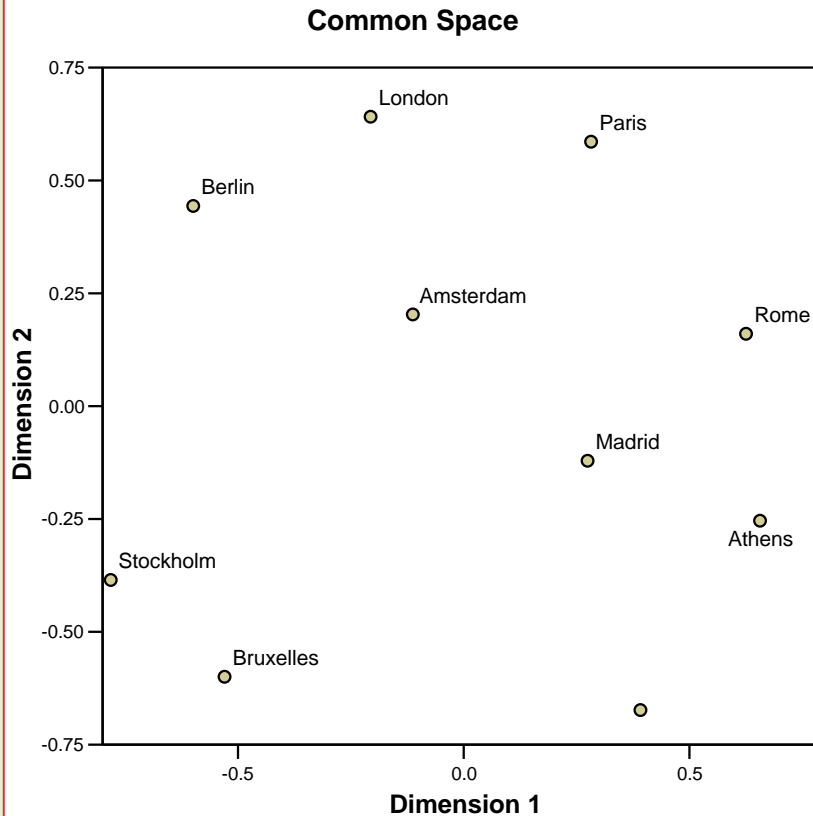
$$d_{ij} = ||\mathbf{x}_i - \mathbf{x}_j|| = \left[\sum_{a=1}^k (x_{ia} - x_{ja})^2 \right]^{1/2} \approx \delta_{ij}$$

Euclidean distance

Boyut Küçültme

Çok Boyutlu Ölçekleme

■ Örnek çıktı



Yorum: Trend olması

- Anket sonucu her kişi şehirleri sıralar.
- Sıralama cinsinden benzerlikler hesaplanır (Londra Atina'ya kıyasla Berlin'e daha çok benzer).
- Eğer bulunan yeni boyutlar yorumlanabiliyorsa, algılanan farklılıklar tespit edilebilir.

İklim olarak düşünülebilir

Boyut Küçültme

Çok Boyutlu Ölçekleme

▣ Fransız şehirleri arası tren zamanları

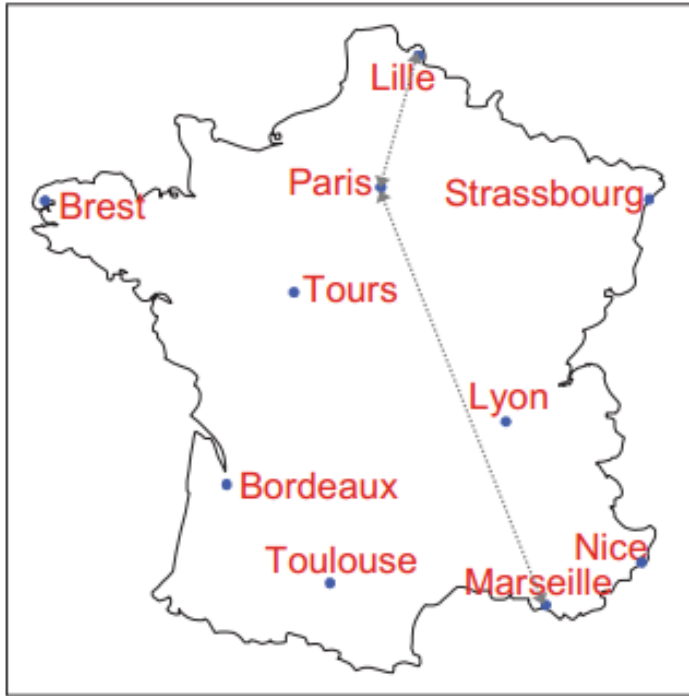
	Bor- deaux	Brest	Lille	Lyon	Mar- seille	Nice	Paris	Strassb ourg	Tou- louse	Tours
Bordeaux	0									
Brest	9:58	0								
Lille	6:39	7:11	0							
Lyon	8:05	7:11	4:52	0						
Marseille	5:47	8:49	6:12	1:35	0					
Nice	8:30	13:36	8:20	4:33	2:26	0				
Paris	2:59	4:17	1:04	2:01	3:00	5:52	0			
Strasbourg	8:08	10:16	6:54	4:36	7:04	11:15	4:01	0		
Toulouse	2:02	13:52	9:42	4:25	3:26	6:29	5:14	10:56	0	
Tours	2:36	5:38	4:17	4:21	5:13	9:04	1:13	6:03	6:06	0

▣ Bu bilgi ile Fransa haritası üzerinde şehirler işaretlenebilir mi?

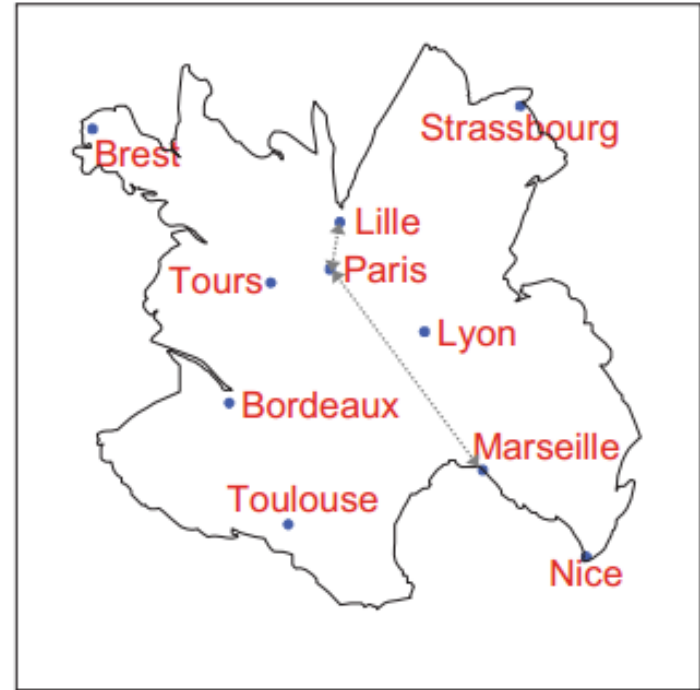
Boyut Küçültme

Çok Boyutlu Ölçekleme

□ Cevap: Evet



Gerçek
Harita



MDS
Harita

Boyut Küçültme

Çok Boyutlu Ölçekleme

▣ R kodları