

# Kümeleme Analizi Sonuçlarının Tematik Haritalar ile Görselleştirilmesi

İbrahim Öztuğ BİLDİRİCİ<sup>1\*</sup>, Nurbanu AFACAN<sup>2</sup>

<sup>1</sup>Selçuk Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Bölümü, Konya  
(bildirici@selcuk.edu.tr) ORCID ID 0000-0001-7717-586X

<sup>2</sup>Bingöl Kadastro Müdürlüğü, Bingöl  
(nurbanuafacan@gmail.com) ORCID ID 0000-0001-7472-6312

## Özet

Kümeleme analizi ile birden çok konuda ortak özellikleri olan nesnelere bulunur. Tematik kartografyada iller, ilçeler gibi veri toplama birimleri nesnelere olarak kabul edilip, kümeleme analizi ile seçilen konularda ortak özelliğe sahip bölgeler belirlenerek, elde edilen sonuçlar tematik haritalar halinde ilgili kullanıcılara sunulabilir. Bu çalışmada bu yaklaşımın uygulanabilirliği ele alınmış, iller bazında seçilen suç ve işsizlik konuları bakımından ortak özelliğe sahip il grupları belirlenmiştir. Bu şekilde ülke bazında seçilen konular açısından görsel bir değerlendirme yapma olanağı karar vericilere sunulmuş olmaktadır.

**Anahtar Kelimeler:** Kümeleme Analizi, Tematik Kartografya, Ward Yöntemi

## Visualization of Clustering Analysis Results with Thematic Maps

### Abstract

In clustering analysis objects with similar characteristics can be identified in terms of two or more selected topics. In thematic cartography enumeration units can be assumed objects, and clustering analysis can be performed with selected data. The results can be visualized in terms of thematic maps. In this study the applicability of this approach is focused on. An application is given on the base of Turkish provinces, the enumeration units here, with two variables: unemployment and crime. So the regions with common properties can be brought into view. The results can be used by decision makers.

**Keywords:** Clustering Analysis, Thematic Cartography, Ward Method

## 1. GİRİŞ

Kümeleme analizi, nesnelere ya da bireyleri benzerliklerine göre kümeler ayırmak için kullanılan çok değişkenli bir istatistik analiz yöntemidir (Tatlıdıl, 1996). Kümeleme analizi sonucu oluşturulan kümeler içinde, aynı küme içinde yer alan nesnelere birbirlerine diğer kümelerde yer alan nesnelere daha çok benzeşirler (Atbaş, 2008). Kümeleme analizi, idari birimlerde (veri toplama birimleri, enumeration unit) toplanmış istatistiksel verilere uygulanarak benzer özellikleri olan bölgeler belirlenip sonuçlar tematik haritalar

biçiminde ilgili kullanıcılar ya da karar vericilere sunulabilir. Bu şekilde tematik haritalar ile belirlenen grupların ya da kümelerin ilgili kullanıcı grubu tarafından anlaşılması kolayca sağlanır. Kümeleme analizi klasik tematik kartografya ders kitaplarında genel olarak yer almamasına rağmen Slocum vd. (2004) konuyu bir örnek uygulama eşliğinde ele almıştır. Selvi ve Çağlar (2016) ve Selvi ve Çağlar (2017) kümeleme analizinde k-means ve k-medoids yöntemlerini uygulayarak elde ettikleri sonuçları tematik haritalar halinde sunmuşlardır.

\* Sorumlu Yazar

Bu çalışmada kümeleme analizinin, Türkiye İstatistik Kurumu verileri ile iller bazında seçilen konularda uygulanması hedeflenmiştir. Bu amaçla Hiyerarşik kümeleme tekniklerinden olan Ward yöntemi seçilerek, Türkiye illerine ait suç ve işsiz sayıları verileri ile uygulama yapılmıştır. Veriler Türkiye İstatistik Kurumu (TÜİK) web sitesinden indirilmiştir (<http://www.tuik.gov.tr>). Çalışma sonucunda işsizlik ve suç işleme olayları açısından ortak özelliklere sahip bölgelerin ortaya konması hedeflenmiştir. Kümeleme analizinin en önemli hususlarından biri küme sayısının belirlenmesidir. Bu nedenle uygulamada üç farklı küme sayısı için ayrı ayrı analiz yapılmıştır. Burada amaç küme sayısı seçiminin sonuçlara etkisini irdelemektir.

Yürütülen çalışmanın ilk sonuçları Afacan ve Bildirici (2016)'da bildiri olarak sunulmuş olup çalışma halen devam etmektedir.

İkinci bölümde kümeleme yöntemi ele alınmış, üçüncü bölümde üç farklı küme sayısı için uygulamalar verilmiş, son bölümde elde edilen sonuçlar ve yakın gelecekte planlanan çalışmalar tartışılmıştır.

## 2. YÖNTEM

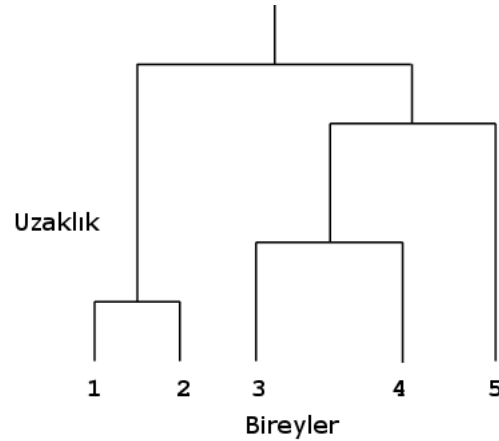
Kümeleme analizinin temel amacı, toplanan çok sayıdaki gruplanmamış gözlemlerden oluşan verilerin, birimlerin benzerliklerine göre anlamlı gruplardan oluşan özel alt kümeler bölünerek verinin indirgenmesini sağlamaktır. En az bilgi kaybıyla, anlaşılabilir tanımlı verilere sahip olunması amaçlanmaktadır (Atbaş, 2008).

Kümeleme analizi ilk kez 1939 yılında Tryon tarafından uygulanmıştır. 1960'lı yıllardan itibaren kullanımı yaygınlaşmıştır. Kümeleme teknikleri, araştırma problemlerinde geniş bir alanda uygulanmaktadır (Anderberg, 1973; Hartigan, 1975).

Birimlerin benzerliklerine göre kümeler ayırmada yararlanılabilecek çeşitli yaklaşımlar vardır. Bu yaklaşımlardan biri, en çok benzer iki birimi aynı gruba atamakla başlayıp tüm birimlerin aynı gruba atanması ile biten hiyerarşik bir yaklaşımdır. Bir diğer yaklaşım ise tüm verilerin ortalama değerlerine en yakın değerlere sahip nesnelere aynı kümeye atanmasını temel alan yaklaşımdır (Blashfield ve Aldenferder, 1978). Yaygın kullanılan bu

iki yaklaşımdan başka yaklaşımlar da vardır. Tüm yaklaşımlarda en önemli ölçüt, kümeler arası farklar ve küme içi benzerliklerin en fazla olmasının sağlanmasıdır.

Hiyerarşik bir yöntem olan Ward yönteminde, grup bağlantılarından çok grup içi kareler toplamı işlenmektedir (Chatfield ve Colins, 1980). Yönteme her birinin içinde tek bir birim bulunan n tane küme ile başlanır. Yöntemin ilk basamağında her gözlem bir küme olduğundan hata kareler toplamı sıfır olmaktadır (Everitt, 1974). Her aşamada iki alt küme bir sonraki seviyeyi oluşturmak için birleştirilir. Birimler değişik seviyelerde başarılı bir şekilde bir araya geldikleri "dendogram" adlı şemada gösterilir (Dibb, 1998, Şekil 1). Ward yöntemi aykırı noktalara duyarlı bir yöntem olduğundan uygulamada kullanılmasına karar verilmiştir (Everitt vd., 2011).



Şekil 1: Dendogram örneği

Kümeleme analizinin uygulama adımları aşağıdaki gibidir.

- Değişkenlerin seçilmesi ve veri matrisinin belirlenmesi
- Birimlerin birbirleriyle olan benzerlik ya da uzaklıklarını gösterecek uygun bir benzerlik/uzaklık ölçütü ile benzerlik/uzaklık matrisinin oluşturulması
- Uygun bir kümeleme tekniği ile benzerlik/uzaklık matrisine göre birimlerin uygun sayıda kümeler ayrılması ve oluşturulan bu kümelerin yorumlanması
- Kümelerin yapılarının kurulan hipotezlerle test edilmesi ve gerekli analitik yöntemlerin uygulanması

### 3. UYGULAMA

Günümüzde işsizliğin suça teşvik eden bir faktör olabileceği değerlendirilmektedir. Bu genel düşüncenin geçerliliğini irdelemek amacıyla il bazında veriler seçilerek Türkiye bazında bir analiz yapılarak sonuçlarının tematik harita biçiminde sunulması hedeflenmiştir.

Kümeleme analizinin ilk aşaması veri kümesinin elde edilmesidir. Bu amaçla ilgili veriler TÜİK web sitesinden indirilmiş ve düzenlenmiştir.

Kümeleme analizinde gerekli hallerde veri kümesi standartlaştırılır. Suçlar; öldürme, yaralama, dolandırıcılık, hırsızlık, yağma, sahtecilik, zimmet, rüşvet vb olmak üzere değişik kategorilerde verilmektedir. Her ile ait suç işleyenlerin toplam sayısı ve işsiz sayısı nüfusa oranlanarak yüzdelik değerler elde edilerek veriler standartlaştırılmıştır.

Hiyerarşik kümeleme analizinde her gözlem ayrı bir küme olarak düşünülür. Daha sonra ilk benzerlik katsayılarıyla birbirine en çok benzeyen kümeler birleştirilir. Kümeleme işlemi, Ward yöntemi kullanılarak Matlab yazılımında gerçekleştirilmiştir. Burada uygun küme sayısının belirlenmesi önemlidir. Uygun sayıda küme seçmek için yaygın olan bir yaklaşım, dendrogramdaki kırılmalara, bir diğeri ise kümenin birleştirilmesi için benzerlik katsayısında oldukça büyük bir sıçramanın olduğu yerlere bakılmasıdır. Bu çalışmada küme sayıları 5, 6 ve 7 olarak seçilerek uygulamalar yapılmıştır. 5, 6 ve 7 küme seçimi ile işsizlik ve suç oranları bakımından ortak özelliklere sahip iller Tablo 1,2 ve 3'de verilmiştir.

Elde edilen kümeler MapInfo yazılımı ile tematik harita biçiminde görselleştirilmiştir.

Türkiye'nin coğrafi konumuna uygun olacak şekilde tematik harita uygulanmasında ölçek 1:8 000 000, projeksiyon ED50 datumunda Lambert Konform Konik (standart paraleller 40°40', 43°20') olarak seçilmiştir. Lambert Projeksiyonu hakkında bilgi için Uçar vd. (2011)'den yararlanılabilir. Şekil 2,3 ve 4'de 5,6 ve 7 küme uygulamasına göre oluşturulan tematik harita örnekleri görülmektedir. Küme sayılarının bu şekilde seçiminin nedenlerinden biri de, tematik haritalarda 5-7 değişik renk

kullanımının renklerin birbirinden kolay ayırt edilmesi açısından en uygun sonucu vermesidir (Slocum vd., 2004). Renk seçimleri çevrimiçi ColorBrewer yazılımı ile yapılmıştır (<http://colorbrewer2.org>).

Her bir kümenin kendi içerisinde ortalama suç ve işsizlik değerleri Tablo 1, 2 ve 3'de verilmiştir. Hesaplanan bu değerler kümelerin suç ve işsizlik değişkenleri açısından karşılaştırmasını kolaylaştırmaktadır. Örneğin 5 küme için 5. küme en düşük suç ve en yüksek işsizlik oranına sahipken 2. küme en yüksek suç ve en düşük işsizlik oranına sahiptir. 6 küme uygulamasında; 6. küme en düşük suç ve en yüksek işsizlik oranına sahipken 4. küme en yüksek suç ve en düşük işsizlik oranına sahiptir. Aynı şekilde 7 küme uygulamasında, 7. küme en düşük suç ve en yüksek işsizlik oranına sahipken 5. küme en yüksek suç ve en düşük işsizlik oranına sahiptir. Tüm küme seçeneklerinde Diyarbakır, Mardin, Siirt, Şanlıurfa, Batman, Şırnak illeri aynı küme içinde yer almıştır. Bu bölge çok düşük suç ve çok yüksek işsizlik oranlarıyla diğer illerden her üç uygulamada da belirgin bir şekilde ayrılmıştır. Yapılan analiz, işsizlik ve suç işleme ilişkisi açısından ülke bazında farklılıkları ortaya çıkartmış, ülke için geçerli genel bir ilişki ortaya koymamıştır. Analiz sonuçları çalışmanın başında ortaya konan işsizliğin suça teşvik edebileceğini varsayımını doğrulamamıştır.

### 4. SONUÇLAR ve TARTIŞMA

Bu çalışmada Türkiye'de istatistik verilerin kümeleme analizi ile gruplanması ve seçilen verilere göre ortak özellikleri olan illerin belirlenmesi ele alınmış, işsizlik suç ilişkisi açısından benzer bölgeleri belirlemek üzere bir uygulama yapılmıştır. Burada verilen sonuçlar yürütülen bir yüksek lisans tez çalışmasının ilk bulgularıdır. Tezin bundan sonraki aşamasında hem seçilen verilerin arasındaki korelasyon hesaplanarak verilerin kümelemeye uygunluğu irdelenecek hem de en uygun küme sayısına ulaşabilmek için testler yapılacaktır. Yüzdelik oranlar olarak gerçekleştirilen veri standartlaştırılmasında da alternatif yaklaşımlar değerlendirilecektir.

## **KAYNAKÇA**

- Afacan, N., Bildirici, İ.Ö. (2017). Tematik Kartografyada Kümeleme Analizi, TMMOB Harita ve Kadastro Mühendisleri Odası, 16. Türkiye Harita Bilimsel ve Teknik Kurultayı, 3-6 Mayıs 2017, Ankara.
- Anderberg, M. R. (1973) Cluster Analysis for Applications, Academic Press, New York.
- Atbaş, A.C.G. (2008). Kümeleme Analizinde Küme Sayısının Belirlenmesi Üzerine Bir Çalışma, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara.
- Blashfield, R.K., Aldenderfer, M.S. (1978) The literature on cluster analysis, *Multivariate Behavioral Research*,13, 271-295
- Chatfield, C., Collins, A.J. (1980) Introduction to Multivariate Analysis, Chapman and Hall, London.
- Dibb, S., (1998) Market Segmentation: Strategies for Success, *Marketing Intelligence & Planning*, 16/7, 394-406.
- Everitt, B., (1974) Cluster Analysis, Heinmann, London.
- Everitt, B., Landau, S., Leese, M., Stahl, D. (2011) Cluster Analysis, 5th Edition, John Wiley & Sons, Chichester.
- Hartigan, J.A. (1975) Clustering Algorithms, Wiley, New York, <http://statsoft.com/textbook/esc.html> [Ziyaret Tarihi: 28.04.2016].
- Selvi, H. Z., Çağlar, B. (2016) Using K-Means and K-Medoids Methods for Multivariate Mapping, *International Journal of Applied Mathematics, Electronics and Computers*, 4, 342-345.
- Selvi, H. Z., Çağlar, B. (2017) Çok Değişkenli Haritalama için Kümeleme Yöntemlerinin Kullanılması, *Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 6(2), 415-429.
- Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2004) Thematic Cartography and Geographic Visualization, 2nd edition, Pearson, USA.
- Tatlıdil, H. (1996) Uygulamalı çok değişkenli istatistiksel analiz, Hacettepe Taş. Yayınları, Ankara.
- Uçar, D., İpbüker, C., Bildirici, İ.Ö. (2011) Matematiksel Kartografya, 2. Baskı, Nobel Yayıncılık, Ankara.

**EKLER**

**Tablo 1. 5 Küme seçiminde oluşan gruplar**

Küme	Grup	Eleman Sayısı	Suç Or. Ort. (%)	İşsizlik.Or.Ort. (%)
1	Adana, Mersin, İzmir, Osmaniye	4	0.29	13.75
2	Antalya, Aydın, Balıkesir, Burdur, Çanakkale, Denizli, Edirne, Manisa, Muğla, Nevşehir, Niğde, Uşak, Karaman, Ardahan	14	0.33	6.29
3	Adıyaman, Ankara, Bitlis, Hakkari, Hatay, İstanbul, Kocaeli, Kahramanmaraş, Muş, Sivas, Van, Yalova	12	0.14	10.70
4	Afyonkarahisar, Ağrı, Amasya, Artvin, Bilecik, Bingöl, Bolu, Bursa, Çankırı, Çorum, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Giresun, Gümüşhane, Isparta, Kars, Kastamonu, Kayseri, Kırklareli, Kırşehir, Konya, Kütahya, Malatya, Ordu, Rize, Sakarya, Samsun, Sinop, Tekirdağ, Tokat, Trabzon, Tunceli, Yozgat, Zonguldak, Aksaray, Bayburt, Kırıkkale, Bartın, Iğdır, Karabük, Kilis, Düzce,	45	0.22	7.15
5	Diyarbakır, Mardin, Siirt, Şanlıurfa, Batman, Şırnak	6	0.11	19.93

**Tablo 2. 6 Küme seçiminde oluşan gruplar**

Küme	Grup	Eleman Sayısı	Suç Or. Ort. (%)	İşsizlik.Or.Ort. (%)
1	Ağrı, Artvin, Bingöl, Bursa, Çankırı, Erzincan, Erzurum, Kars, Kırşehir, Malatya, Rize, Tekirdağ, Trabzon, Yozgat, Aksaray, Bayburt, Iğdır	17	0.19	6.95
2	Afyonkarahisar, Amasya, Bilecik, Bolu, Çorum, Elazığ, Eskişehir, Gaziantep, Giresun, Gümüşhane, Isparta, Kastamonu, Kayseri, Kırklareli, Konya, Kütahya, Ordu, Sakarya, Samsun, Sinop, Tokat, Tunceli, Zonguldak, Kırıkkale, Bartın, Karabük, Kilis, Düzce	28	0.24	6.49
3	Adana, Mersin, İzmir, Osmaniye	4	0.29	13.75
4	Antalya, Aydın, Balıkesir, Burdur, Çanakkale, Denizli, Edirne, Manisa, Muğla, Nevşehir, Niğde, Uşak, Karaman, Ardahan	14	0.31	6.29
5	Adıyaman, Ankara, Bitlis, Hakkari, Hatay, İstanbul, Kocaeli, Kahramanmaraş, Muş, Sivas, Van, Yalova	12	0.14	10.70
6	Diyarbakır, Mardin, Siirt, Şanlıurfa, Batman, Şırnak	6	0.11	19.93

Tablo 3. 7 Küme seçiminde oluşan gruplar

Küme	Grup	Eleman Sayısı	Suç Or. Ort. (%)	İşsizlik.Or.Ort. (%)
1	Bolu Isparta Kayseri Sakarya Zonguldak Düzce	6	0.27	8.98
2	Afyonkarahisar, Amasya, Bilecik, Çorum, Elazığ, Eskişehir, Gaziantep, Giresun, Gümüşhane, Kastamonu, Kırklareli, Konya, Kütahya, Ordu, Samsun, Sinop, Tokat, Tunceli Kırıkkale, Bartın, Karabük, Kilis	22	0.24	6.93
3	Ağrı, Artvin, Bingöl, Bursa, Çankırı, Erzincan, Erzurum, Kars, Kırşehir, Malatya, Rize, Tekirdağ, Trabzon, Yozgat, Aksaray, Bayburt, Iğdır	17	0.19	6.95
4	Adana, Mersin, İzmir, Osmaniye	4	0.29	13.75
5	Antalya, Aydın, Balıkesir, Burdur, Çanakkale, Denizli, Edirne, Manisa, Muğla, Nevşehir, Niğde, Uşak, Karaman, Ardahan	14	0.31	6.29
6	Adıyaman, Ankara, Bitlis, Hakkari, Hatay, İstanbul, Kocaeli, Kahramanmaraş, Muş, Sivas, Van, Yalova	12	0.14	10.70
7	Diyarbakır, Mardin, Siirt, Şanlıurfa, Batman, Şırnak	6	0.11	19.93



Şekil 2. 5 küme seçildiğinde oluşan tematik harita



Şekil 3. 6 küme seçildiğinde oluşan tematik harita





Şekil 4. 7 küme seçildiğinde oluşan tematik harita