



Mayıs 2024

GRAFİKSEL VERİ ANALİZİ

**Yığılmış Alan Grafiği (Stacked Area Chart)
İncelenmesi ve Uygulanması**

Dr. Öğr. Üyesi Onur TOKA

Ece KANLI
2200329048

1 Yığılmış Alan Grafiği

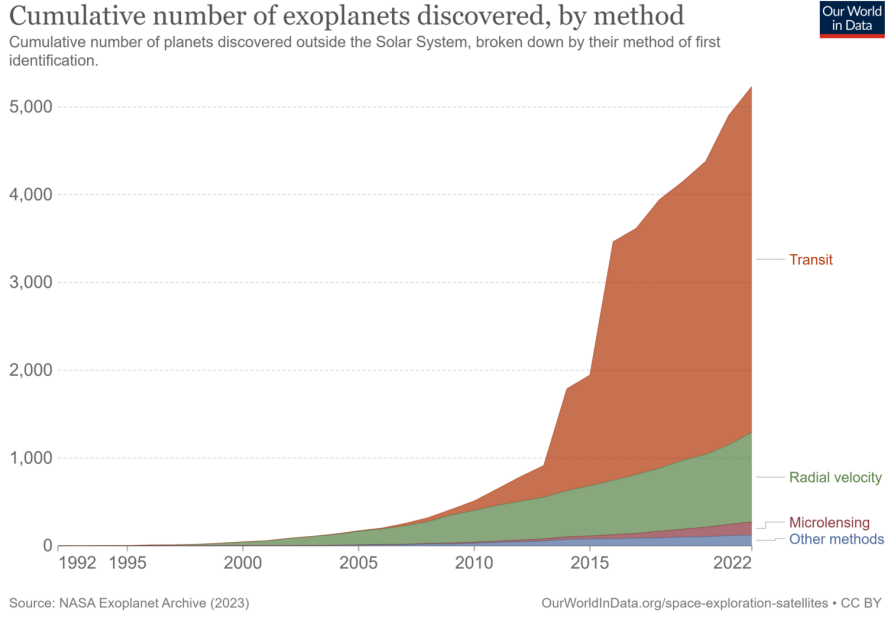
1.1 Tanımı ve Yorumlanması

Yığılmış alan grafikleri, birden fazla kategorinin sayısal değerlerinin genellikle zaman içinde nasıl değiştiğini gösteren bir grafik türüdür. Çizgi ve çubuk grafiklerini birleştirerek altta kalan alanı renklendirilerek belirtilir[1]. Her biri alan olarak belirtilen kategori serileri birbirlerinin üzerine yığılır. En alttaki ilk seri taban çizgisi olarak x eksenini kullanılır. Ancak sonraki her seri, hemen altındaki serinin değerini taban çizgisi olarak kullanır. En üstteki eğri, tüm kategorilerin toplamına karşılık gelirken, altındaki her eğri, altındaki tüm kategorilerin kümülatif katkısını gösterir. Sonuç olarak bu grafik, toplamaları ile farklı kategori değerlerinden gelen katkıları görselleştirir[2][3].

Yığılmış alan grafikleri, parçanın bütüne katkısı vurgulanmak istendiğinde, kategori değerlerinden gelen katkıların zaman içindeki değişimleri gösterilmek istendiğinde ve rakamlar hakkında kaba bir fikir verilmek istendiğinde kullanılır[3].

Örneğin:

Şekil 1’de 1992-2022 yılları arasında 4 yöntemle keşfedilen dış gezegenlerin kümülatif sayısını veren yığılmış alan grafiği verilmiştir.



Şekil 1: Dış Gezegenlerin Keşfi: Yığılmış Alan Grafiği

Grafik incelendiğinde,

- Son on yılda keşfedilen dış gezegenlerin toplam sayısındaki büyük artış görülmektedir. Ayrıca dört yöntem ile de dış gezegen keşfi artış gösterirken özellikle ‘transit’ yöntemi ile keşfedilen dış gezegen sayısında büyük bir artış söz konusudur. Bu durum parça-bütün ilişkisi hakkında bir bakış açısı sunar.
- 2006 yılından önce ‘transit’ yöntemi kullanılarak neredeyse hiç dış gezegenin keşfedilmediğine ve bu dönemdeki keşiflerin çoğunun radyal hız yöntemi

kullanılarak yapıldığı, 2013 ile 2014 yılları arasında ani bir artış ile ‘transit’ yöntemi yoluyla yapılan dış gezegen keşiflerin sayısında kademeli bir artış görüldüğü ve bundan sonraki yıllarda da dış gezegen keşfinde birincil yöntem haline geldiği söylenebilir. Bu durumda kategori değerlerinden gelen katkıların zaman içindeki değişimleri hakkında bir bakış açısı sunar.

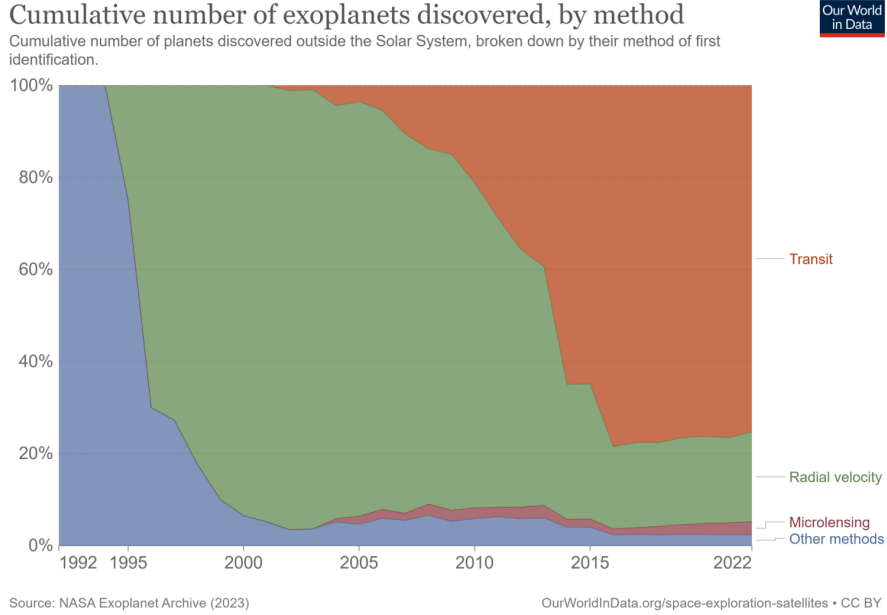
- Grafikte de görüldüğü üzere, yöntemler aynı yatay taban çizgisinde yani x ekseninde yer almıyor ve üst üste diziliyor. Bu durum yöntemleri karşılaştırmayı zorlaştıran bir durumdur. Bu nedenle bu grafiklerin belirli bir kategorinin tam değeri yerine toplamı ve katkıları hızlı bir şekilde göstermek için kullanmanın tavsiye edildiği anlamına gelir[3].

1.2 %100 Yığılmış Alan Grafiği

Yığılmış alan grafiğinin bir çeşidi %100 yığınlı alan grafiğidir.

Standart yığılmış alan grafikleri, toplamın ve kategorilerin mutlak değerleri önemli olduğunda en iyi şekilde kullanılırken %100 yığılmış alan grafikleri, x ekseninde zamana karşı y ekseninde mutlak sayılar yerine yüzde katkısını gösterir. Bu durumda odak noktası mutlak rakamlardan ziyade her bir kategorinin yaptığı katkının oranıdır. %100 yığılmış alan grafikleri, yüzde katkısının daha yüksek düzeyde anlamlı olduğu durumlarda kullanılır.

Örneğin:

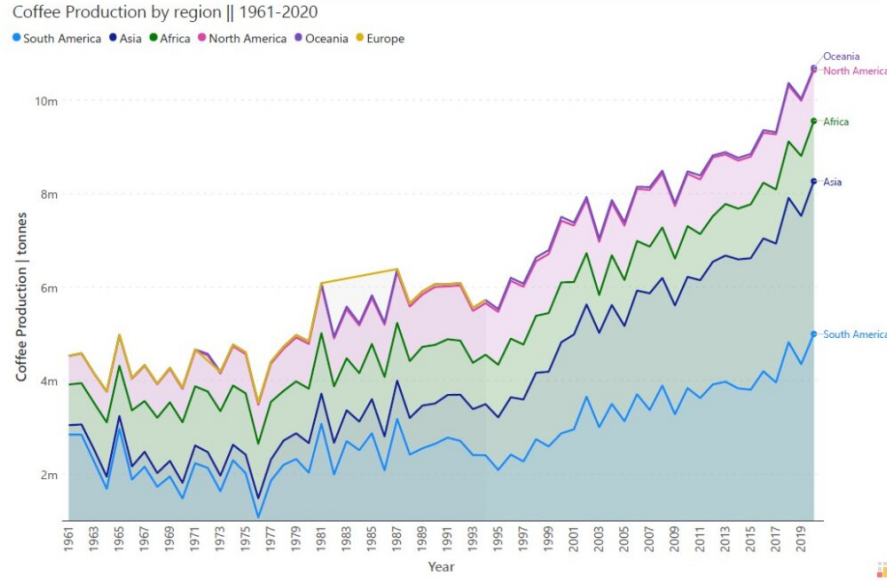


Şekil 2: Dış Gezegenlerin Keşfi: %100 Yığılmış Alan Grafiği

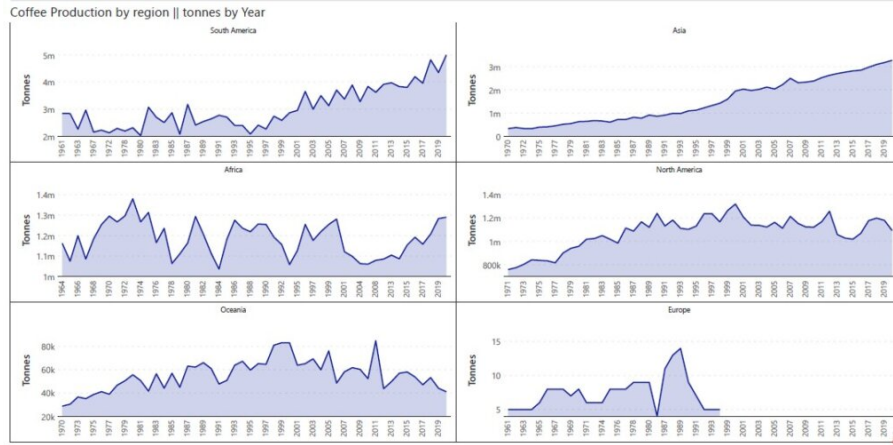
Şekil 2’de, keşfedilen dış gezegenlerin kümülatif sayısına ilişkin aynı veriler, %100 yığılmış alan grafiği olarak çizilmiştir. Bu veri için %100 yığılmış alan grafiği, standart yığılmış alan grafiğine göre daha açıklayıcı olduğu görülmektedir. İlk yıllarda dış gezegen keşfinde diğer yöntemler, son yıllarda ‘transit’ yönteminin baskın olduğu ancak daha çok ‘radial velocity’ yöntemi kullanımının katkı sağladığını söyleyebiliriz.[2].

1.3 Kullanılmaması Gereken Durumlar

1. Her kategorideki dalgalanmalar doğru bir şekilde gösterilmek istediğinde:



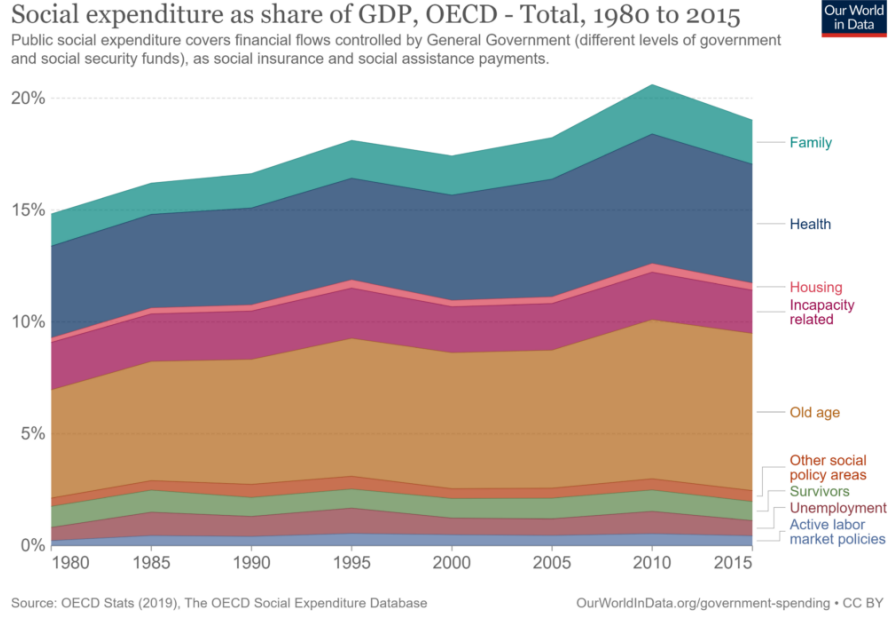
Şekil 3: Ülkelere Göre Kahve Üretimi: Yığılmış Alan Grafiği



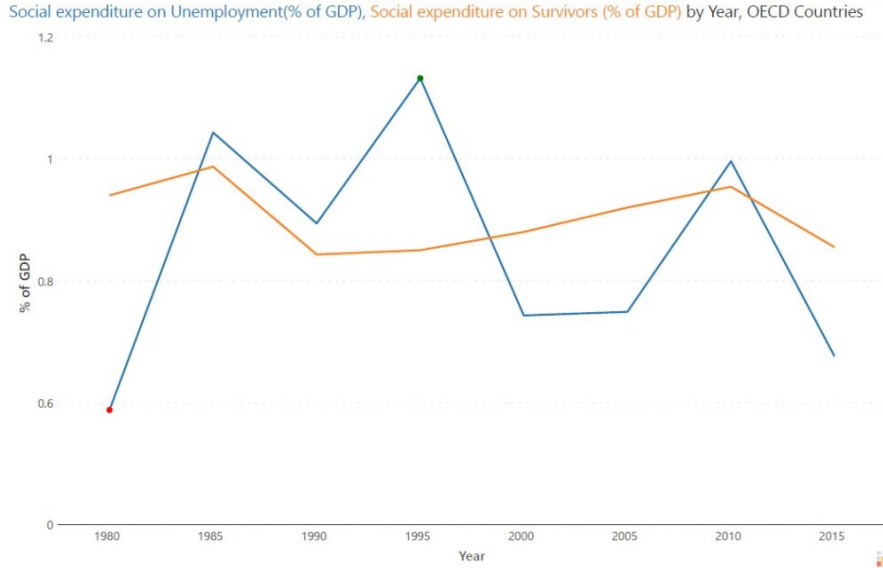
Şekil 4: Ülkelere Göre Kahve Üretimi: Ülkelere Ait Alan Grafikleri

Şekil 3 ve Şekil 4'te verilen grafiklerde 1961-2020 yılları arasında bölgelere göre kahve üretimi verilmiştir. Şekil 3'te verilen grafik, yığılmış alan grafiğidir. Bu grafik ile genel bir bakış açısı yorumu yapılırken her bir ülkeye ait kahve üretimi hakkında daha detaylı yorum yapmak için Şekil 4'te verilen grafiğin kullanımı daha uygun olacaktır. Çünkü bu grafikte tüm kategoriler tek bir yatay temele yerleştirildiğinden zaman içindeki değişimleri daha net görülebilmektedir[3].

2. Bir kategorinin diğerini geride bırakması durumu:



Şekil 5: Sosyal Harcamalar: Yığılmış Alan Grafiği



Şekil 6: Sosyal Harcamalar: Sektör Çizgi Grafiği

Şekil 5'te verilen grafikte 1980-2015 yılları arasında OECD ülkelerinde sosyal harcamaların sektörler üzerinde GSYİH'in yüzdesi olarak gösterildiği yığılmış alan grafiği verilmiştir. Bu harcamaların zaman içerisindeki değişimi hakkında ve hükümetin yıl bazında ne kadar sosyal harcama yaptığına ilişkin yorum yapılabilir. Ancak, hükümetin işsizlikle ilgili harcamaları, hayatta kalanlara yönelik harcamalarla karşılaştırılması yapılmak istenirse yığılmış alan grafiğinden bu yorumu çıkaramayız. Bu durumda karşılaştırılması yapılmak

istenen kategorilerin Şekil 6’da verilen grafikte olduğu gibi bir çizgi grafiği çizdirirsek sektörlerin(kategorilerin), yıllara ve birbirlerine göre değişimleri net bir şekilde görebiliriz[3].

2 Veri Seti Üzerinde Uygulanması

Yığılmış alan grafiği, Kaggle üzerinden alınan [4] ve [5] veri setlerinin birleştirilmesi ile oluşturulan yeni veri seti üzerinde uygulanmıştır.

2.1 Veri Seti Hakkında

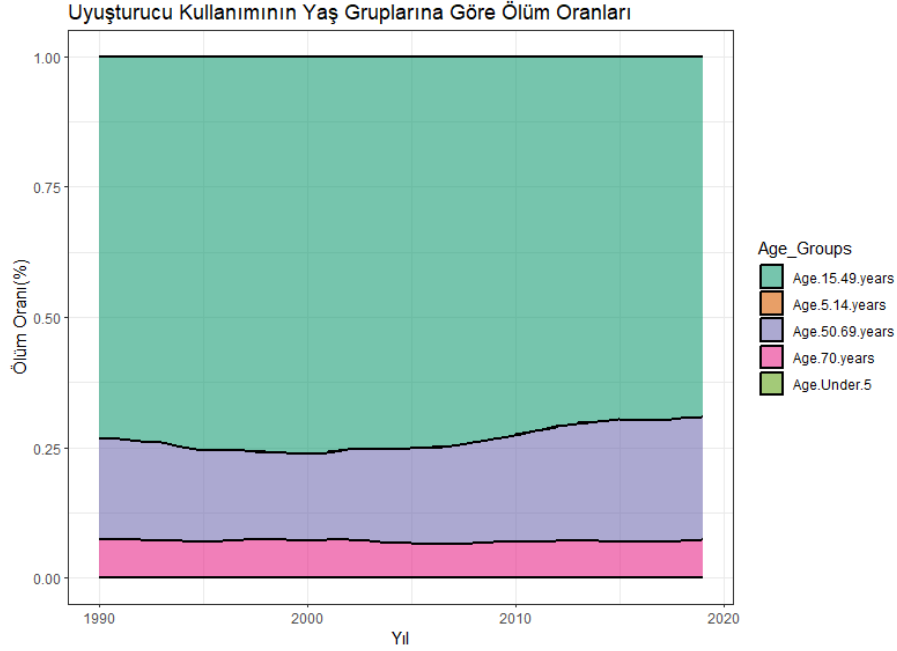
Kullanılan veri seti, 1990-2019 yılları arasında ülke bazında yaş gruplarına ve farklı uyuşturucu madde kullanımına bağlı olarak gerçekleşen ölüm sayıları ile ilgilidir. Değişkenler:

- **Country:** Ülkeleri, bölgeleri ve gelir düzeyine göre ülkeleri içeren değişken. (veri tipi: string)
- **Year:** 1990-2019 yıllarını içeren değişken. (veri tipi: date)
- **Age.Under.5, Age.70.years, Age.15.49.years, Age.50.69.years, Age.5.14.years:** Uyuşturucu kullanımına bağlı olarak yaş gruplarına göre ölüm sayılarını veren değişkenler. (veri tipi: integer)
- **Amphetamine, Cocaine, Opioid, Other.drugs:** Kullanılan uyuşturucuya göre tüm yaş gruplarında ölüm sayılarını veren değişkenler. (veri tipi: integer)

Veri seti, toplamda 6840 gözlem ve 11 değişkenden oluşmaktadır.

2.2 Veri Görselleştirme

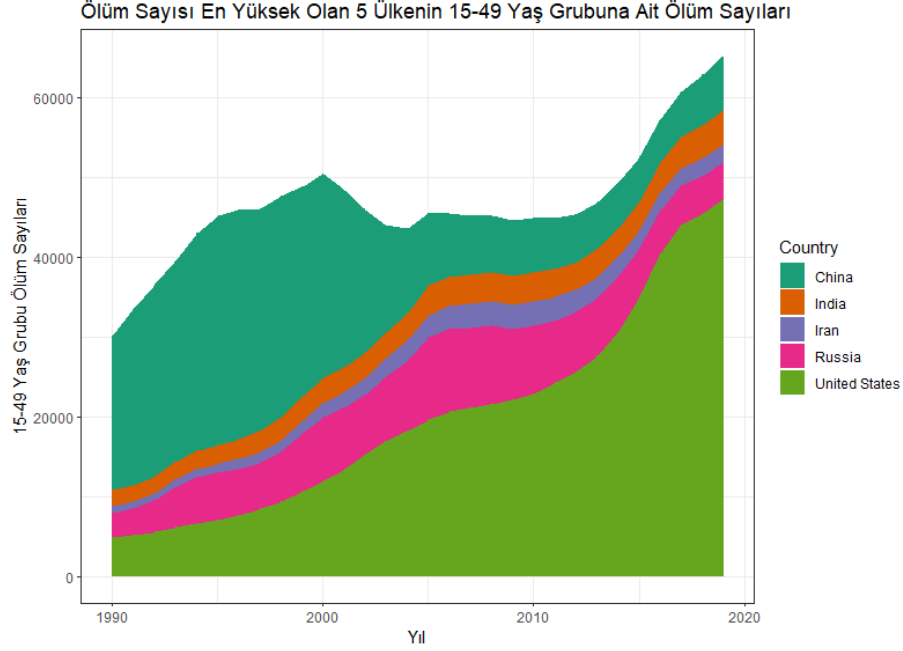
2.2.1 Uyuşturucu Kullanımının Yaş Gruplarına Göre Ölüm Oranları



Şekil 7: Uyuşturucu Kullanımının Yaş Gruplarına Göre Ölüm Oranları

Tüm ülkelerde uyuşturucu kullanımının yaş gruplarına göre ölüm oranını veren %100 yığılmış alan grafiği Şekil 7’de verilmiştir. Grafiğe göre, 1990-2019 yılları arasında tüm ülkelerde uyuşturucu kullanımı sonucunda ölüm oranı en yüksek olan yaş grubu 15-49’ dur. Bu yaş grubunu sırasıyla 50-69 ve 70+ yaş takip etmektedir. 5 yaş altı ve 5-14 yaş grubunda uyuşturucu kullanımına bağlı olarak belirgin bir ölüm oranı gözlenememektedir. Ayrıca 2005-2015 yılları arasında 50-69 yaş grubunda ölüm oranında artış görülmektedir.

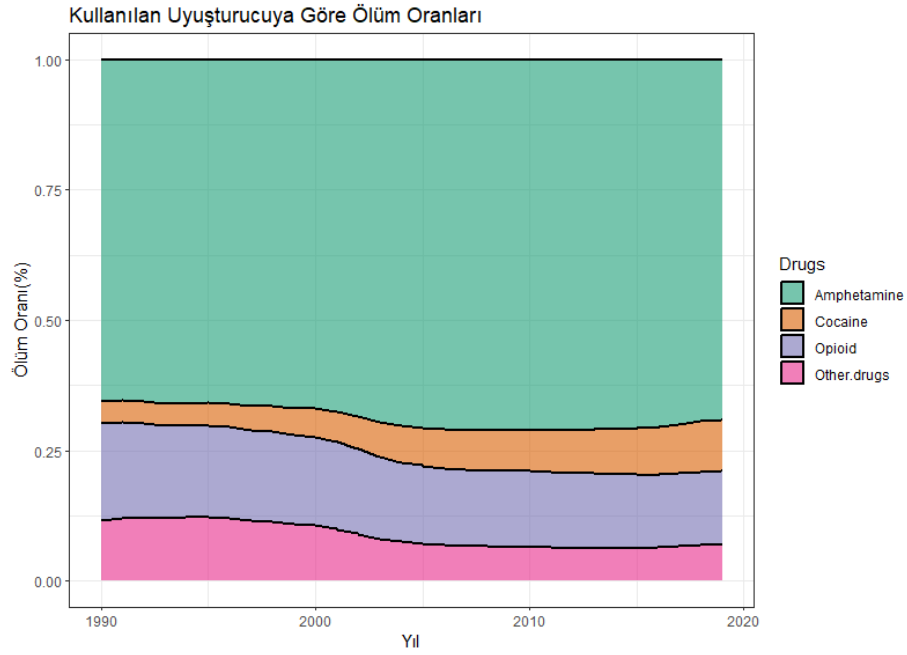
2.2.2 Ölüm Sayısı En Yüksek Olan 5 Ülkenin 15-49 Yaş Grubuna Ait Ölüm Sayıları



Şekil 8: Ölüm Sayısı En Yüksek Olan 5 Ülkenin 15-49 Yaş Grubuna Ait Ölüm Sayıları

Uyuşturucu kullanımında ölüm sayısı en yüksek 5 ülkenin: Çin, Hindistan, İran, Rusya ve Amerika'nın, 15-49 yaş grubuna ait ölüm sayılarını veren yığılmış alan grafiği Şekil 8'de verilmiştir. Grafiğe göre, 2003-2004 yılına kadar Çin'de 15-49 yaş grubunda uyuşturucu kullanımına bağlı çok fazla ölüm gerçekleştiği ve sonrasında ölüm sayılarının azaldığı, Amerika'da 2010 yılı civarı dışında yıllar geçtikçe artan ölümlerin olduğu, Rusya, Hindistan ve İran'da ise 2005-2010 yılları arasında artan ölümlerin olduğu söylenebilir.

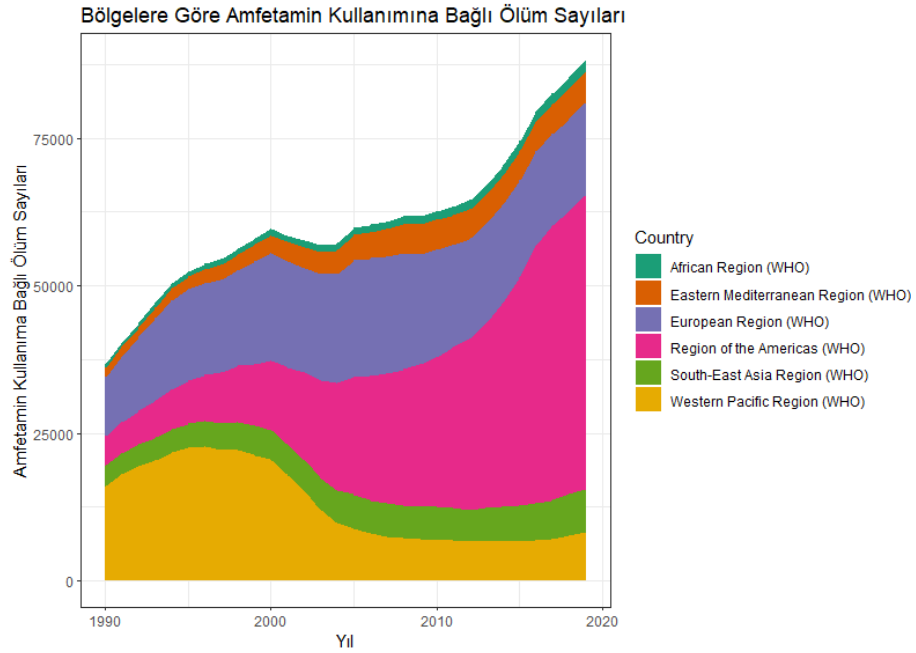
2.2.3 Kullanılan Uyuşturucuya Göre Ölüm Oranları



Şekil 9: Kullanılan Uyuşturucuya Göre Ölüm Oranları

Tüm ülkelerde, kullanılan uyuşturucuya göre ölüm oranlarını veren %100 yığılmış alan grafiği Şekil 9'da verilmiştir. En fazla ölüme neden olan uyuşturucu, amfetamindir. Kokaine bağlı ölüm oranı, yıllar geçtikçe artmaktadır. Opioid ve diğer uyuşturucu madde kullanımında ölüm oranının 2000 yılından sonra azalmaya başladığı ve neredeyse sabit bir ölüm oranı gözlemlendiği söylenebilir.

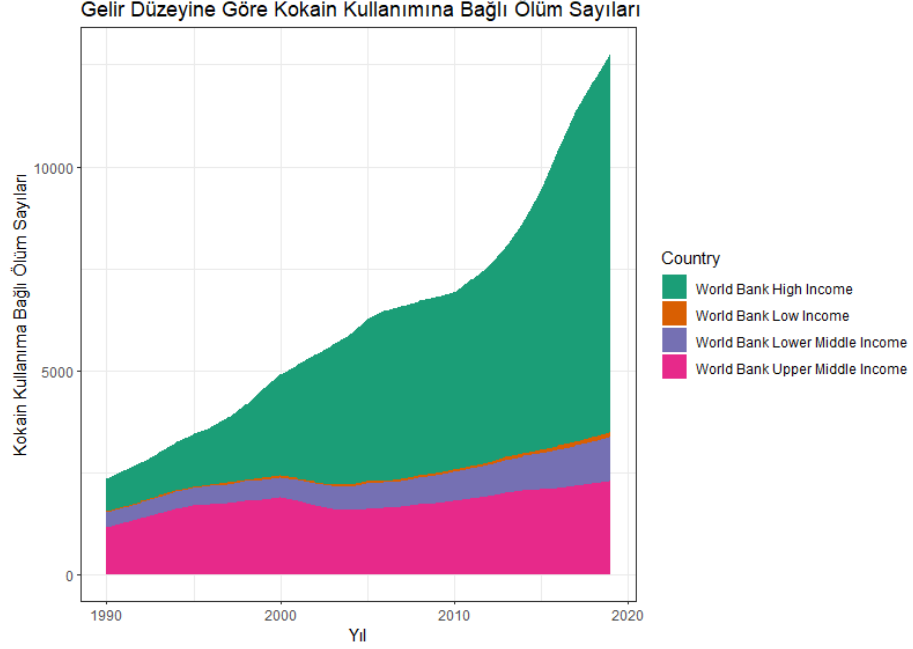
2.2.4 Bölgelere Göre Amfetamin Kullanımına Bağlı Ölüm Sayıları



Şekil 10: Bölgelere Göre Amfetamin Kullanımına Bağlı Ölüm Sayıları

Bölgelere göre amfetamin kullanımına bağlı ölüm sayılarının yığılmış alan grafiği Şekil 10'da verilmiştir. Amerika bölgesinde, 2000 yılından sonra amfetamine bağlı ölüm sayısında ciddi artışlar gözlenmektedir. Batı Pasifik bölgesinde 2000 yılına kadar amfetamine bağlı ölüm sayıları fazlayken 2000 yılından sonra daha az ölümler meydana gelmiştir. Avrupa bölgesinde 2000-2010 yılları arasında ölüm sayıları fazladır.

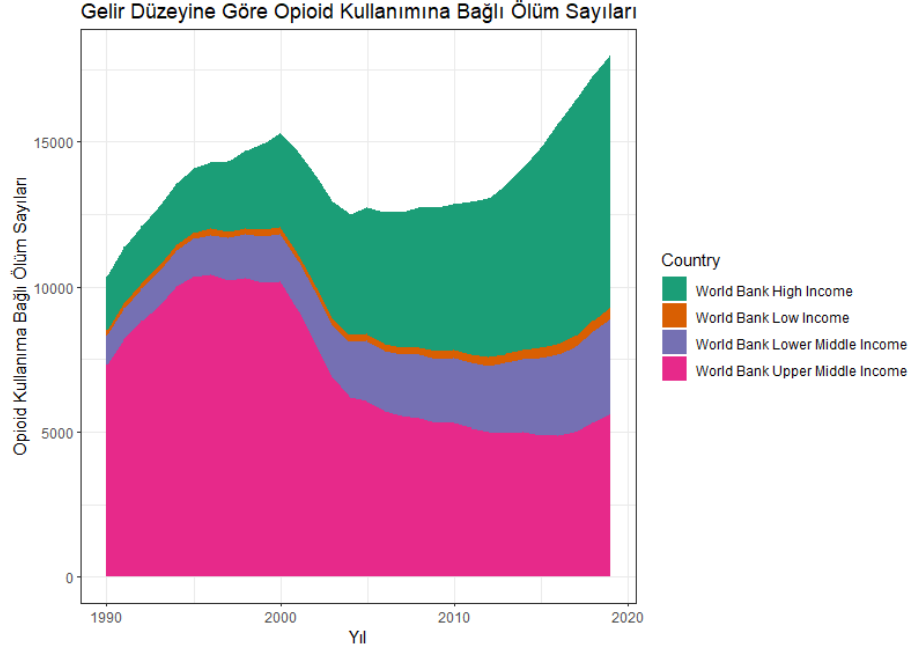
2.2.5 Ülkelerin Gelir Düzeyine Göre Kokain Kullanımına Bağlı Ölüm Sayıları



Şekil 11: Ülkelerin Gelir Düzeyine Göre Kokain Kullanımına Bağlı Ölüm Sayıları

Ülkelerin gelir düzeyine göre kokain kullanımına bağlı ölüm sayılarının yığılmış alan grafiği Şekil 11’de verilmiştir. Gelir düzeyi yüksek olan ülkelerde kokaine bağlı ölüm sayısı oldukça yüksektir. Bu sayının son yıllarda özellikle arttığı söylenebilir. Kokaine bağlı ölüm sayısının yüksek olduğu bir diğer gelir grubu ise üst-orta gelire sahip ülkelerdir. Özellikle 2000 yılı ve 2010 yılından sonra ölüm sayılarında artış gözlemlenmektedir.

2.2.6 Ülkelerin Gelir Düzeyine Göre Opioid Kullanımına Bağlı Ölüm Sayıları



Şekil 12: Ülkelerin Gelir Düzeyine Göre Opioid Kullanımına Bağlı Ölüm Sayıları

Ülkelerin gelir düzeyine göre opioid kullanımına bağlı ölüm sayılarının yığılmış alan grafiği Şekil 12’de verilmiştir. Gelir düzeyi üst-orta olan ülkelerde opioide bağlı ölüm sayısı 1995-2000 yılları arasında oldukça yüksektir. Bu sayı son yıllarda azalsada ölüm sayısının hala yüksek olduğu söylenebilir. Yüksek gelirli ülkelerde ise 2005 yılından sonra , düşük-orta gelirli ülkelerde 2010 yılından sonra ölüm sayısının arttığı görülmektedir.

3 Kodların İncelenmesi

```
1. ggplot(graph1, aes(x=Year, y=percentage, fill=Age_Groups)) +  
  geom_area(alpha=0.6 , size=1, colour="black") +  
  scale_fill_brewer(palette = "Dark2") +  
  xlab("Yıl") +  
  ylab("Ölüm Oranı(%)") +  
  ggtitle("Uyuşturucu Kullanımının Yaş Gruplarına Göre Ölüm Oranları")
```

Bu kod satırı (Şekil 7):

- **ggplot()** fonksiyonu ile **graph1** veri seti kullanılarak yıllara göre, yaş gruplarına göre yüzde
- **aes()** fonksiyonu ile yıllara (**Year**) göre yaş gruplarının (**Age_Groups**) yüzde değerlerine (**percentage**) ilişkin grafiği belirtir.
- **geom_area()** fonksiyonu, alan grafiğini oluşturur. **alpha=0.6** ile alanların saydamlığı belirlenirken, **size=1** ile çizgi kalınlığı ve **colour="black"** ile çizgi rengi ayarlanır.
- **scale_fill_brewer()** fonksiyonu, renk paletini belirler. "Dark2" paleti kullanılarak yaş gruplarının her birine farklı renkler atanır.
- **xlab()**, **ylab()** ve **ggtitle()** fonksiyonları eksen etiketlerini ve grafiğin başlığını belirtir.

```
2. ggplot(graph2, aes(x=Year, y=Age.15.49.years, fill=Country)) +  
  geom_area() + scale_fill_brewer(palette = "Dark2") +  
  xlab("Yıl") +  
  ylab("15-49 Yaş Grubu Ölüm Sayıları") +  
  ggtitle("Ölüm Sayısı En Yüksek Olan 5 Ülkenin 15-49 Yaş  
  Grubuna Ait Ölüm Sayıları")
```

Bu kod satırı (Şekil 8):

- **ggplot()** fonksiyonu ile **graph2** veri seti kullanılarak yıllara göre, yaş gruplarına göre yüzde
- **aes()** fonksiyonu ile yıllara (**Year**) göre ülkelerin (**Country**) 15-49 yaş grubuna (**Age.15.49.years**) ilişkin grafiği belirtir.

Kaynaklar

- [1] Atlassian, “A complete guide to area charts.” <https://www.atlassian.com/data/charts/area-chart-complete-guide> [Accessed: (25.04.2024)].
- [2] Inforiver, “Stacked area charts - an overviews.” <https://inforiver.com/insights/stacked-area-charts-an-overview/> [Accessed: (25.04.2024)].
- [3] Inforiver, “Stacked area charts: When to use them and when to avoid them?” <https://inforiver.com/insights/stacked-area-charts-when-to-use-and-avoid/> [Accessed: (25.04.2024)].
- [4] Kaggle, “Cocaine, opioids, cannabis and illicit drugs.” <https://www.kaggle.com/datasets/programmerrdai/cocaineopioidscannabis-and-illicit-drugs?select=deaths-from-drug-use-disorders-by-age.csv> [Accessed: (28.04.2024)].
- [5] Kaggle, “Cocaine, opioids, cannabis and illicit drugs.” <https://www.kaggle.com/datasets/programmerrdai/cocaineopioidscannabis-and-illicit-drugs?select=deaths-drug-overdoses.csv> [Accessed: (28.04.2024)].
- [6] Stackoverflow, “Convert four digit year values to class date.” <https://stackoverflow.com/questions/30255833/convert-four-digit-year-values-to-class-date>[Accessed: (29.04.2024)].
- [7] T. R. G. Gallery, “Stacked area chart with r.” <https://r-graph-gallery.com/136-stacked-area-chart.html>[Accessed: (30.04.2024)].