

VERİ GÖRSELLEŐTİRME ÇALIŐMASI

Hazırlayan: Emre Erol

Pozisyon: Stajyer



İÇİNDEKİLER

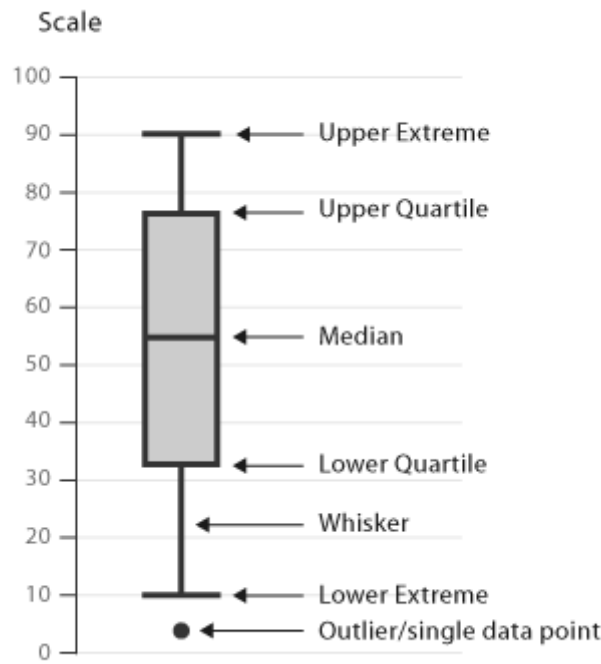
VERİ GÖRSELLEŞTİRME ÇALIŞMASI	1
BOX PLOT.....	4
Box Plot Nedir?	4
Şekil 1	4
Box Plot'un Avantajları	4
Box Plot'un Dezavantajları	4
Box Plot'un Kullanım Alanları	5
Şekil 2.....	6
VIOLIN GRAPH.....	6
Violin Graph Nedir?	6
Şekil 3.....	7
Violin Graph'ın Avantajları	7
Violin Graph'ın Dezavantajları	7
Violin Graph'ın Kullanım Alanları	8
Şekil 4.....	8
LINE GRAPH	8
Line Graph Nedir?	8
Şekil 5.....	9
Line Graph'ın Avantajları.....	9
Line Graph'ın Dezavantajları	9
Line Graph'ın Kullanım Alanları	9
Şekil 6.....	10
BAR GRAPH	10
Bar Graph Nedir?	10
Şekil 7	11
Bar Graph'ın Avantajları.....	11
Bar Graph'ın Dezavantajları	11
Bar Graph'ın Kullanım Alanları	11
Şekil 8.....	12

PIE CHART	12
Pie Chart Nedir?.....	12
Şekil 9.....	13
Pie Chart'ın Avantajları	13
Pie Chart'ın Dezavantajları.....	13
Pie Chart'ın Kullanım Alanları.....	13
Şekil 10	14

BOX PLOT

Box Plot Nedir?

Şekil 1'de de görüldüğü üzere box plot grafiği veri setinin alt çeyrek, medyan ve üst çeyrek değerlerini alıp bunlar özelinde veri setinin gösterilmesini ve tanımlanmasını hedefler. Whisker yani bıyık olarak çevirebileceğimiz kısımlar ile çeyrekler arasındaki kısımları görece fazla veya az fakat uç derecede olmayacak şekilde nitelendirebiliriz. Upper Extreme (Üst Aşırı) veya Lower Extreme (Alt Aşırı) sınırları geçildiği takdirde sınırları geçen değerler için outlier (aşırı) tabirini kullanabiliriz. Veri setinde beklenmedik bir değerde olduklarını da söylemek bunu belirtmenin başka bir yolu olabilir.



Şekil 1

Box Plot'un Avantajları

Görsel olarak okunmasının kolay olması yanında bize bir veri setine dair verilerin konumlarını, yayılımını, çarpıklığını ve dağılımın kuyruğu hakkında da bilgiler verir. Veri setinde beklenenden fazla veya az olan değerlerin gösterim olarak nokta şeklinde belirtilmesi de okunmasını kolay ve kullanılmasını verimli hale getiren özelliklerden birisidir. Kutunun ortası olarak medyan değerinin seçilmesi ve kenarlarını belirleyen değerlerin ise alt çeyrek ve üst çeyrek değerler olması şekli veri setinin ortalamasından kurtarır ve outlier yani aşırı değerlere karşı dirençli bir konuma sokar.

Box Plot'un Dezavantajları

Bize her ne kadar öz bir şekil ile zengin bir bilgi havuzu sağlasa da veri üzerinde daha detaylı görselleştirmelerde ve yoğunluk dağılımının detaylı incelenmesi gibi durumlarda yeterli bilgiyi

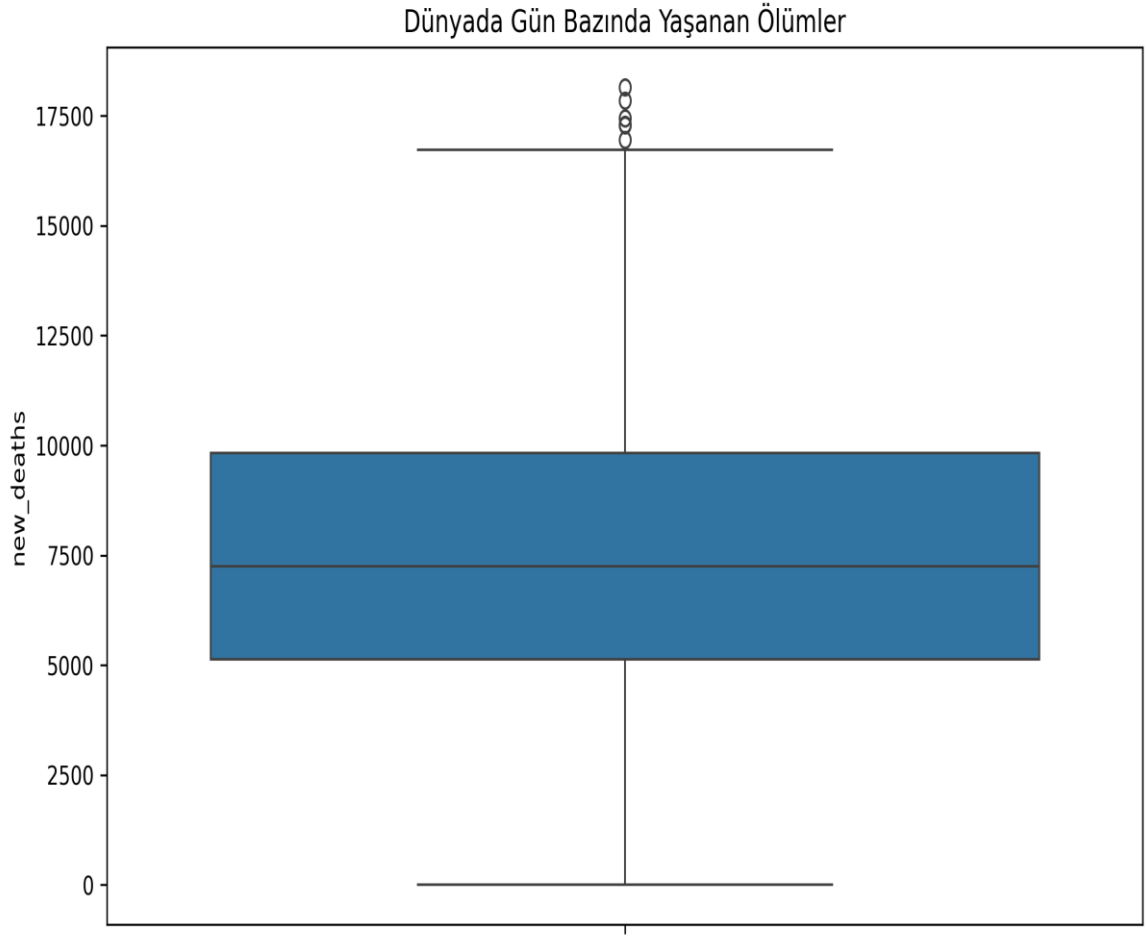
sağlayamamaktadır. Ortalama, standart sapma ve verilerin sayısı gibi konularda da şekil bazında eksik kalmaktadır. Özetlemek gerekirse, daha detay gerektiren konularda box plot pek tercih edilen bir gösterim değildir.

Box Plot'un Kullanım Alanları

Box plot grafiği, sayısal veriler üzerinden ilk bakışların atılmasında ve veriler üzerinde dağılım incelenmesi özelinde aşırı değer tespiti gibi konularda kullanılabilir. Aşırı değerlerin varlığı ve sayıları gibi konular ile birlikte kutunun bulunduğu aralıklar incelenebilir ve incelenen veri üzerinde faydalı yorumlar yapılabilir. **Şekil 2**'de ise Covid-19 virüsü yüzünden günlük olarak yaşanan ölümlerin bir kutu grafiği üzerinde gösterimi mevcut. Görüldüğü üzere her gün yaşanan ölümlerin medyanı 7500 imiş ve 160.000 gibi bir değerden sonrası ise aşırıya kaçan değerler olarak nitelendirilmiş. Bu bilgiler ışığında ilerde new_deaths üzerinden yapılabilecek herhangi bir işlemde önce aşırı değerler üzerinde düşünülmeli ve alınan kararlar neticesinde ilerlenilmelidir. Veri setinin büyüklüğü arttıkça homojenlik çok bozulmadığı takdirde gösterim olarak yine dirençli ve uyumlu bir yapı göstermesi de onu genel hatları aktarımı açısından daha güzel bir potaya taşımaktadır.

- Finans ve Ekonomi
- Tıp ve Biyoloji
- Süreç Analizi
- Büyük Veri Setleri
- Vb.

Gibi alanlarda da yukarıda belirtilen sebeplerden ötürü kullanılmaktadır.



Şekil 2

VIOLIN GRAPH

Violin Graph Nedir?

En basit tabiriyle violin graph, box plot (kutu grafiği) etrafına aynalanmış bir yoğunluk dağılımı üzerine tasarlanmış bir görselleştirme yöntemi olarak tanımlanabilir. [Şekil 3](#)'te de görüldüğü üzere içerisindeki yapı bakımından kutu grafiklerinin yöntemini ve gösterimini içermektedir. Etrafındaki çizgiler ise yoğunluk grafiğimizin sınırlarını temsil etmektedir. Bu yapısı bakımıyla kemanı çağrıştırdığından dolayı literatüre violin graph (keman grafiği) olarak geçmiştir.

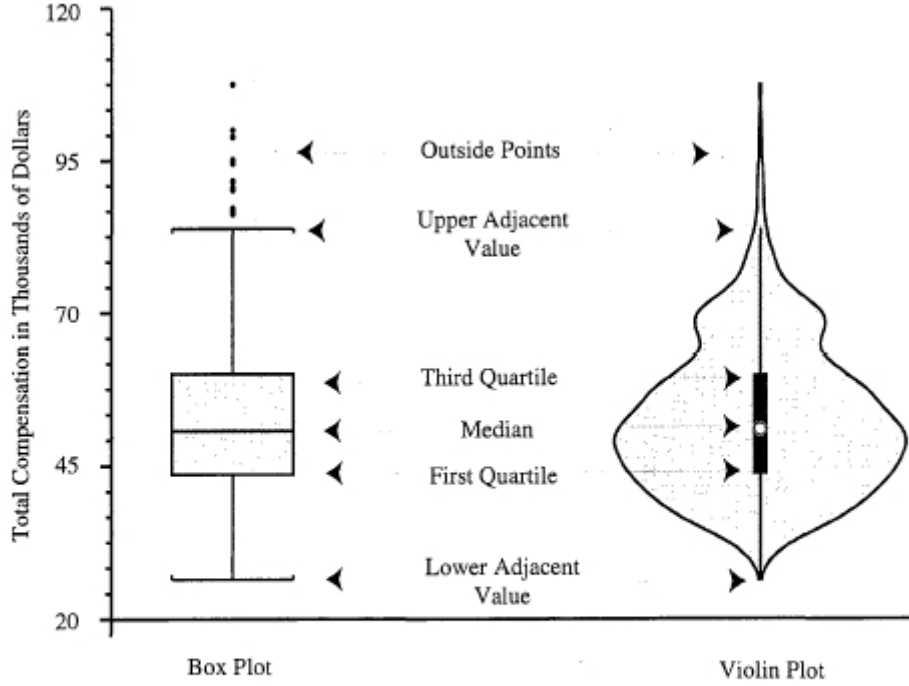


Figure 1. Common Components of Box Plot and Violin Plot. Total compensation for all academic ranks.

Şekil 3

Violin Graph'ın Avantajları

Violin graph'ın avantajlarına box plot özelinde yazdığımız avantajları eklemek herhangi bir sorun yaratmayacaktır. Kendi iç yapısı kutu grafikleri üzerine kurulu olduğundan dolayı kutu grafiklerinin de avantajlarını barındırmaktadır. Bunlar: aykırı değer tespiti, aşırı değerlere karşı dirençli olması, veri setinin yayılımı gibi konular olabilir. Tabii ki kendisini kutu grafiklerinden daha avantajlı konuma sokan yanları da mevcuttur. Çarpıklık gibi konularda veri setinin dağılımını incelemek istediğimiz zaman sınırlı sayıda bilgi veren box plot grafiklerinden ziyade direkt olarak bize yoğunluk dağılımını da gösteren violin plot'u seçmek doğru bir tercih olacaktır. Bu yönüyle bize daha detaylı ve dolayısıyla daha fazla bilgi vermektedir. Ayıryeten boyut arttırımı gibi durumlarda da arttırılan boyut ile birlikte ayrılan yoğunluk dağılımları da karşılaştırma açısından büyük bir avantaj sağlamaktadır.

Violin Graph'ın Dezavantajları

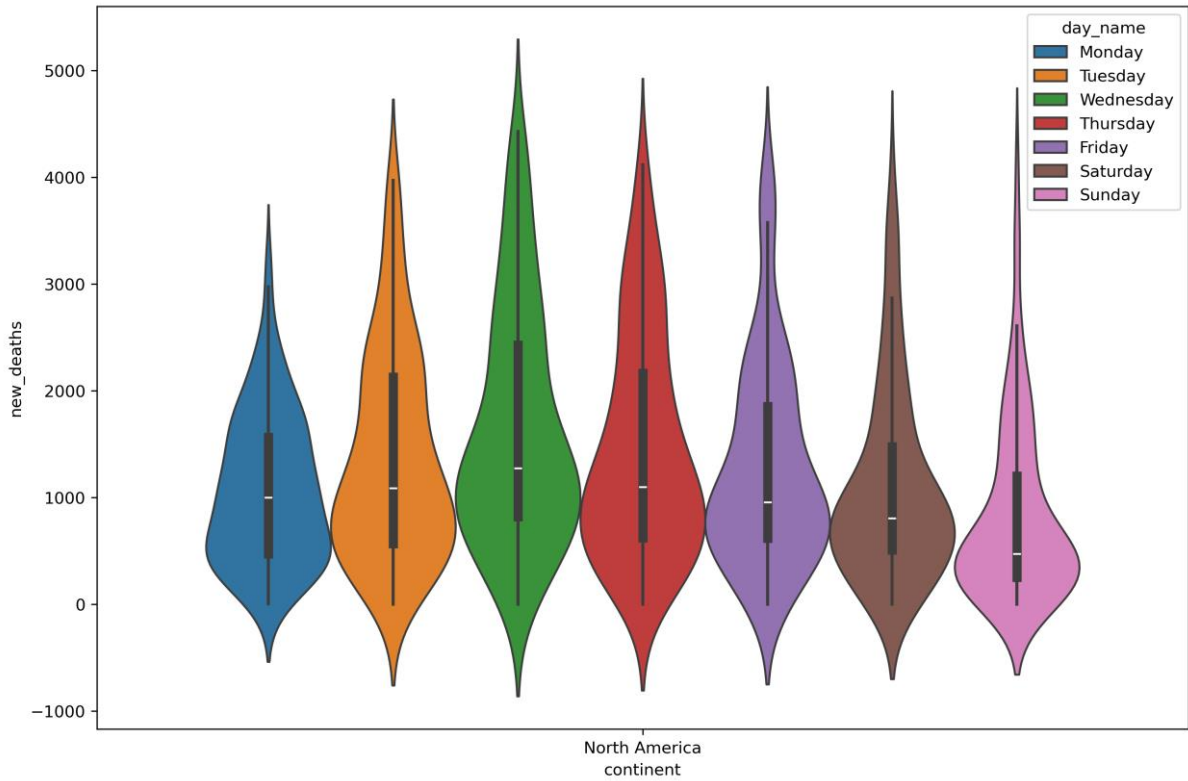
Yukarda bahsettiğimiz bazı avantajların kimi durumlarda getirdiği dezavantajlar da pekâlâ bulunmaktadır. Daha fazla detay ve bilgi verme durumu bazı durumlarda okumasını daha zor ve efor sarf ettiren bir potaya sokabilmektedir. Ayıryeten yoğunluk dağılımı gösteriminden dolayı aykırı değerlerin net bir şekilde görülmesi de kutu grafiklerine nazaran daha zor olmaktadır. İstatistiksel olarak da violin graph'ların o kadar da iyi olmadığını öyle ki violin graph'ların hiçbir işe yaramadığını veya kötü olduklarını savunan insanlar bile mevcuttur. Buradaki temel görüş dağılımın medyan bilgisini geçersiz kıldığı durumlarda kutu grafiğinin bir anlamının kalmaması ve dağılım medyan bilgisini yansıtırken ise dağılım bilgisinin gereksiz bir bilgilendirme olmasından kaynaklı. Elbette bütün bu görüşlere rağmen kendisine literatürde yer edinmiş ve kullanımı devam eden grafiklerden birisidir.

Violin Graph'ın Kullanım Alanları

Violin graph'ların kullanım alanı oldukça geniş olabilmektedir. Veri setlerinin karşılaştırılmasında ve çarpıklık, dağılım ve aykırı değer gözlemlerinde kullanılmaktadır. **Şekil 4**'te görüldüğü üzere North America'da yaşanan Covid-19 virüsü bazlı günlük ölüm sayılarının günlere göre bölünmüş bir gösterim yer almaktadır. Buradaki dağılımlara bakarak medyan bilgisinin anlamlı bir potada olduğunu görebiliyor ve buna bağlı olarak günlerin medyanları özelinde hangi gün daha çok ölüm gerçekleşmiş görebiliyoruz. Her gün türünde yaşanan aşırı ölüm değerleri olduğunu ve en çok da Sunday yani pazar günü yaşandığını keman grafiği yardımıyla görebilmekteyiz. Bu gibi aşırı değerlere rağmen bizim daha sağlıklı bir bilgi edinmemizi sağlayan keman grafiğinin medyan bazlı olması da onu bir adım öne çıkarmaktadır.

- İstatistiksel Analizler
- Finans ve Ekonomi
- Büyük Veri Setleri
- Karşılaştırma Problemleri
- Vb.

Gibi alanlarda da kendisine yer edinmiş ve kullanımı devam etmiştir.



Şekil 4

LINE GRAPH

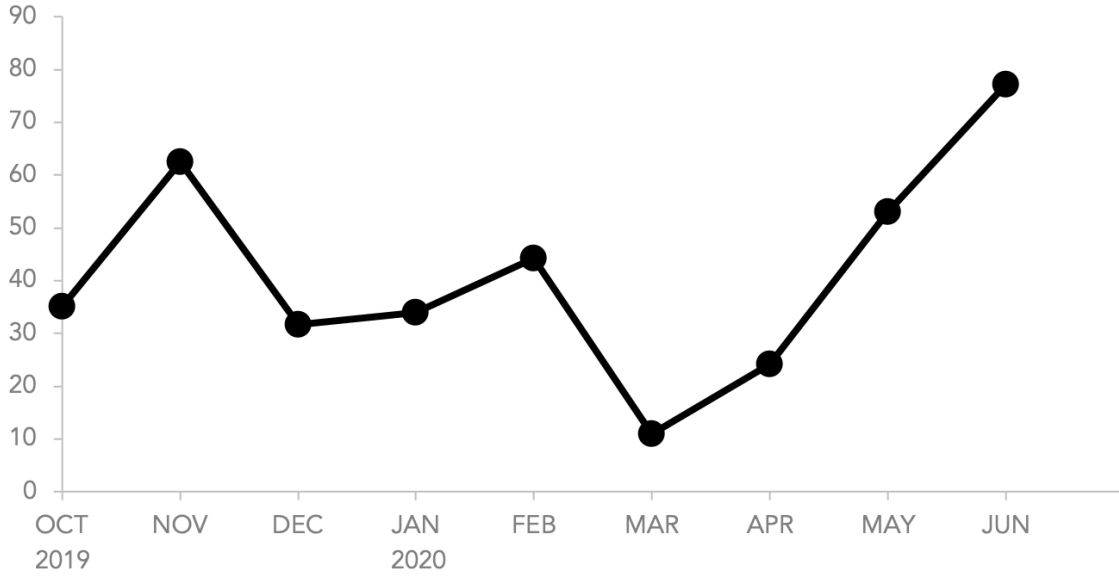
Line Graph Nedir?

Line graph yani çizgi grafikleri temel olarak belirli noktalar arasında çizgiler çekerek veri setinin olağan akışını göstermeyi amaçlar. Bu akış gösterimi özellikle tarih bilgisi içeren veri setlerinde faydalı

olabilmekte ve daha zengin bir bilgi havuzu sunabilmektedir. **Şekil 5**'te de görüldüğü üzere x ekseninde bir zaman bilgisi ve y ekseninde de o zaman bilgisi özelinde gerçekleşen satış bilgileri yer almaktadır. Böylece aylara bölünmüş olan x ekseninin içini çizgiler yardımıyla daha net bir şekilde görebilmekteyiz.

Produce sales

IN THOUSANDS (USD)



Şekil 5

Line Graph'ın Avantajları

Çizgi grafiklerinin birçok avantajı bulunmaktadır. Bunlar: akış içerisindeki veri setlerinin karşılaştırılması, çizgiler yardımıyla daha detaylı bir veri seti gösterimi, zaman eksenini içeren grafiklerde zengin bilgi aktarımı, vb. gibi konular olabilir. Ayrıyeten çizgi içermesinden kaynaklı eğim gibi bilgileri de incelememiz mümkün hale gelir ve prediction (tahmin) ya da forecasting (öngörü) gibi eylemler de daha kolay ve mümkün hale gelmektedir. Basit ve anlaşılır olması da süreçlerin daha verimli geçmesinde avantajlı özelliklerden bir tanesidir.

Line Graph'ın Dezavantajları

Veri setinin içindeki noktaların çok düzensiz olması durumunda arasına çekilen çizgiler görsel olarak okunmasını zorlaştırır ve gerekli bilgilerin sağlanmasında zorluk çıkartır. Dolayısıyla büyük veri setlerinde görselleştirilmesi görece biraz daha zor ve uğraştırıcı bir potadadır. Eğer veri setinin dağılımındaki homojenlik düşük ise yine görsel olarak okuması zor bir çıktı verebilir.

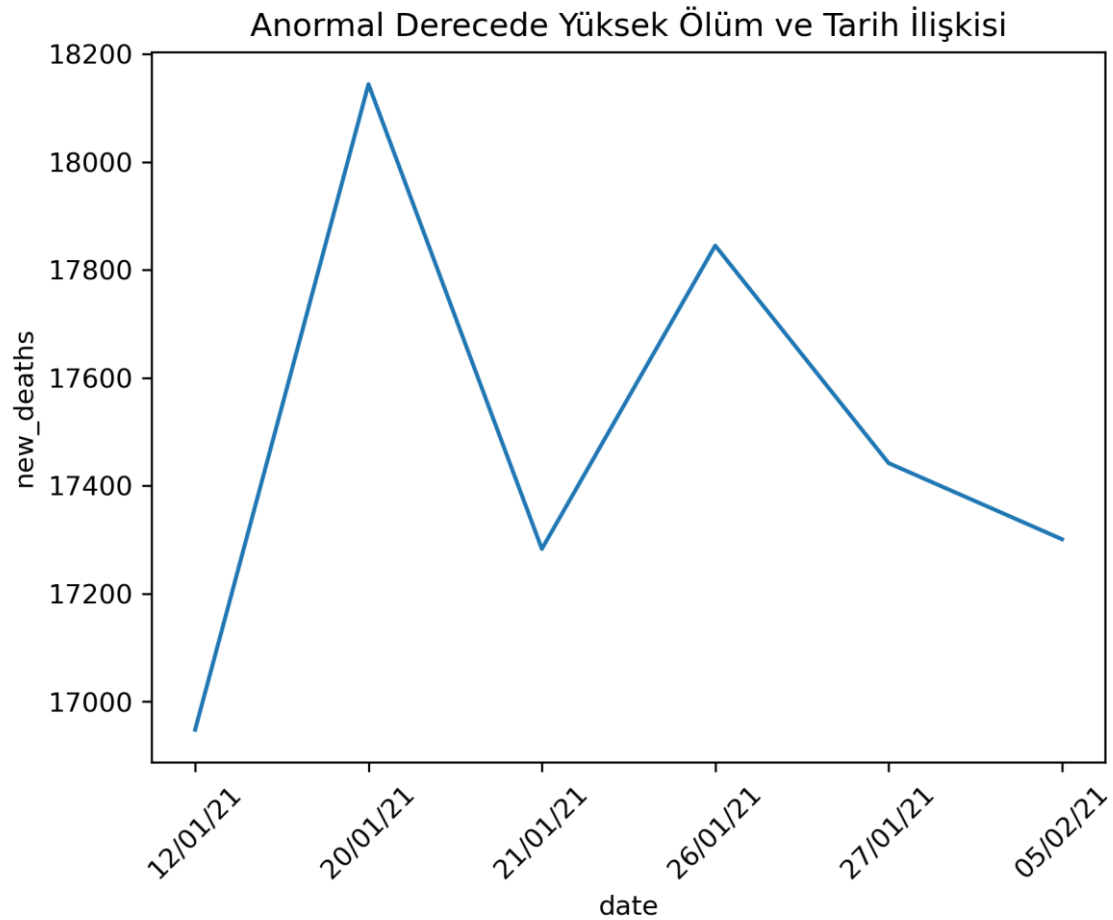
Line Graph'ın Kullanım Alanları

Gündelik hayatta bir veri setinin akışının izlemek adına oldukça faydalı olan çizgi grafiklerinin benzer bir örneğini **Şekil 6**'da görebilmekteyiz. Bu grafik bize daha öncesinde box plot'tan elde ettiğimiz aşırı değer bilgilerinin değerlerini zaman indeksli bir şekilde akış içerisinde göstermektedir. Elimizdeki

bilgiler sadece o tarihlerde olmasına rağmen o günler arasında gerçekleşen ölümlere dair de çizgiler aracılığıyla bir bilgi sunar.

- Zaman Serisi Analizleri
- Trendlerin Görselleştirilmesi
- Değer Karşılaştırmaları
- Bilimsel Araştırmalar ve Deneyler
- Finansal Grafikler
- Vb.

Gibi alanlarda da kullanımı devam etmektedir.

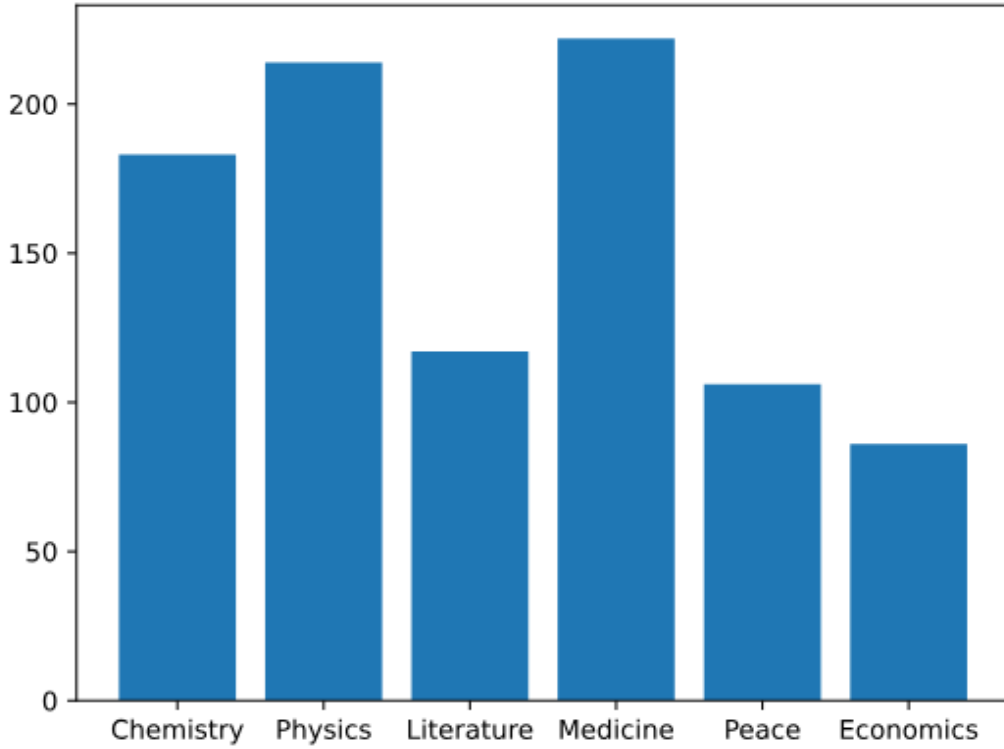


Şekil 6

BAR GRAPH

Bar Graph Nedir?

Bar grafiği isminden de anlaşıldığı üzere ayrı barlardan oluşan ve genellikle kategorik veri ile numerik veri arasındaki ilişkiyi gösteren bir grafik türüdür. Veri setinin dağılımında, kategorik verilerin daha iyi açıklanmasında ve kategorik veriler arasındaki karşılaştırmalarda oldukça faydalı olan bir gösterim türüdür. **Şekil 7**'de de belirtildiği üzere çeşitli başlıklar özelinde frekanslarının dağılımları gösterilmiştir. Bu frekans eksenini farklı bir numerik sütun da olabilirdi ve bize veri seti hakkında daha detaylı bir bilgi verilmesi de mümkün olabilirdi.



Şekil 7

Bar Graph'ın Avantajları

Bar grafiğinin akla gelen ilk avantajlarından birisi kolay okunabilirliği ve barlar aracılığıyla aktarılan ilişkinin kategorik ve numerik olmak üzere iki farklı türü ele almasından kaynaklı daha zengin bir şekilde bilgiyi sunmasıdır. Bize sadece bu iki farklı tür arasındaki ilişkiyi anlatmakla kalmaz ve kimi zaman yoğunluk hakkında da bilgi verebilir. Hem yatay hem de dikey olmak üzere iki farklı şekilde sunulabilir olması da anlatım kolaylığı açısından önemli bir detaydır. Boyut arttırmalarına karşı sunduğu bilginin karmaşıklığı çok artmazken sunulan bilginin zenginliği önemli ölçüde artabilmektedir.

Bar Graph'ın Dezavantajları

Sütun sayılarının fazla olması durumunda aktarılan bilginin kalitesinde düşmeler yaşanabilir ve dağılım net bir şekilde aktarılmayabilir. Bu durum aynı zamanda görsel olarak da okumayı zorlaştırabilir. Aşırı detay bazında da bize gerekli bilgileri yeterli düzeyde sağlayamamaktadır. Bazı durumlarda kategoriler arasındaki farkın az olması gibi durumlarda fark edilmesi zor görseller de sunabilmektedir.

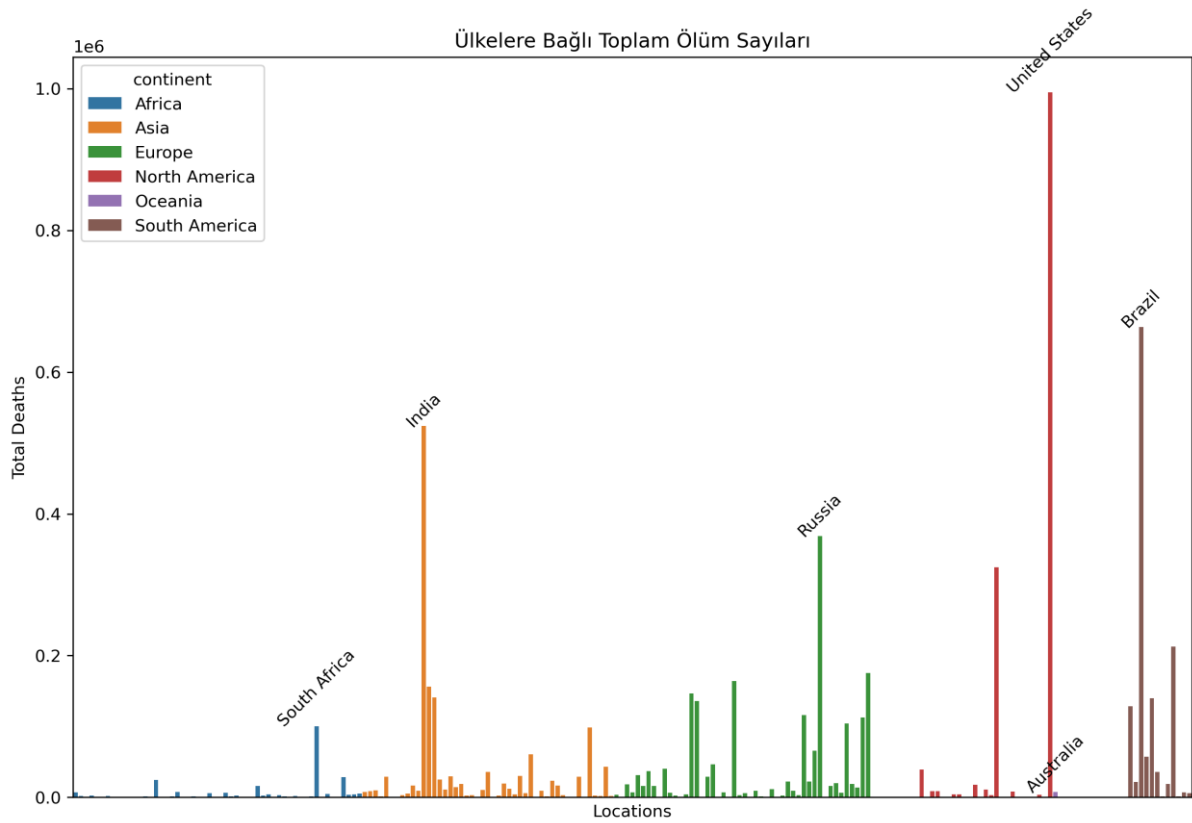
Bar Graph'ın Kullanım Alanları

Bar grafiklerinin kullanım alanı oldukça geniştir. **Şekil 8** aracılığı ile bir örnek ele almış olalım. Görüldüğü üzere bize bu grafik Locations yani ülkeler ile Total Deaths yani Covid-19 virüsü sebebiyle yaşanan toplam ölümler hakkında bilgi vermektedir. Kıtalar ise ekstra bir boyut olarak eklenmiş ve ayrı renklerle kendilerini etiketlemiştir. Kıtalar bazında bakıldığında zaman Afrika kıtasının en çok

ölüm yaşayan ülkesi Güney Afrika, Güney Amerika kıtasının en çok ölüm yaşayan ülkesi Brezilya gibi bilgilere rahatlıkla ulaşılabilir. Bu durum bize bar plot'un gösterim açısından ne kadar rahat ve zengin bir bilgi aktarımı ve havuzu olduğunu hatırlatmaktadır.

- Kategorik Verilerin Karşılaştırılması
- Demografik Verilerin Analizi
- Bilimsel Araştırmalar ve Deneysel Sonuçlar
- Anket ve Kamuoyu Araştırmaları
- Vb.

Gibi alanlarda da kendisine yer edinmiştir ve kullanımı devam etmektedir.

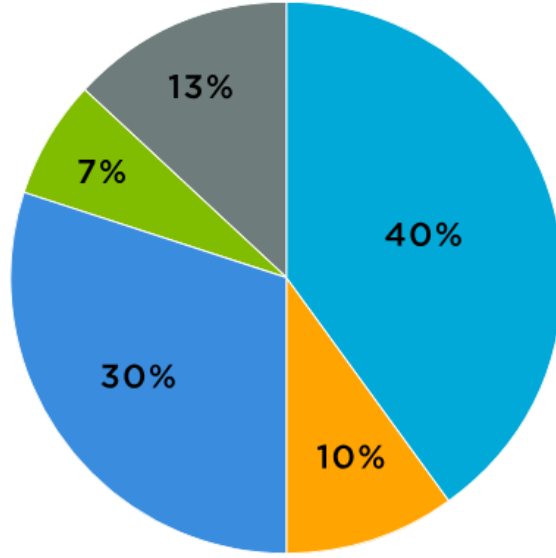


Şekil 8

PIE CHART

Pie Chart Nedir?

Pie chart yani pasta grafiği elimizdeki veri setinin farklı sınıflarını bölüp bir daire içerisinde çeşitli yüzdeliklere aktararak gerekli bilgiyi vermeyi hedefler. Bu yapıyla okumayı oldukça kolaylaştıran ve kıyaslamalar konusunda rahatlık sağlayan pie chart çıkarttığı görsel bakımıyla pastaya benzemesinden bu ismi almıştır. Şekil 9'da da görüldüğü üzere farklı sınıflar farklı renklere atanmıştır ve toplam yüzdelikleri 100 olacak şekilde farklı yüzdeliklere bölünmüşlerdir. Hepsinin toplamının 100 olması hepsinin bir bütünden geldiğini ve o bütün özelinde dağılımlarının kapladıkları yerleri aktardıklarını gösterir.



Şekil 9

Pie Chart'ın Avantajları

Basit bir düzlemde sınıfların yüzdelerik dilimlerinin ortak bir özellik üzerinden verilmesi karşılaştırma problemlerini ve veriden gelen görselleştirmelerin anlaşılma adımlarını büyük ölçüde kolaylaştırmaktadır. Böylelikle verilerin büyüklüğüne dair herhangi bir bilgi edinmeden oranlara dair bilgiler elde edebilir veya verilerin büyüklüklerine dair bilgimiz olsa da aralarındaki farkı görmek açısından daha rahat bir görüş kazanabiliriz.

Pie Chart'ın Dezavantajları

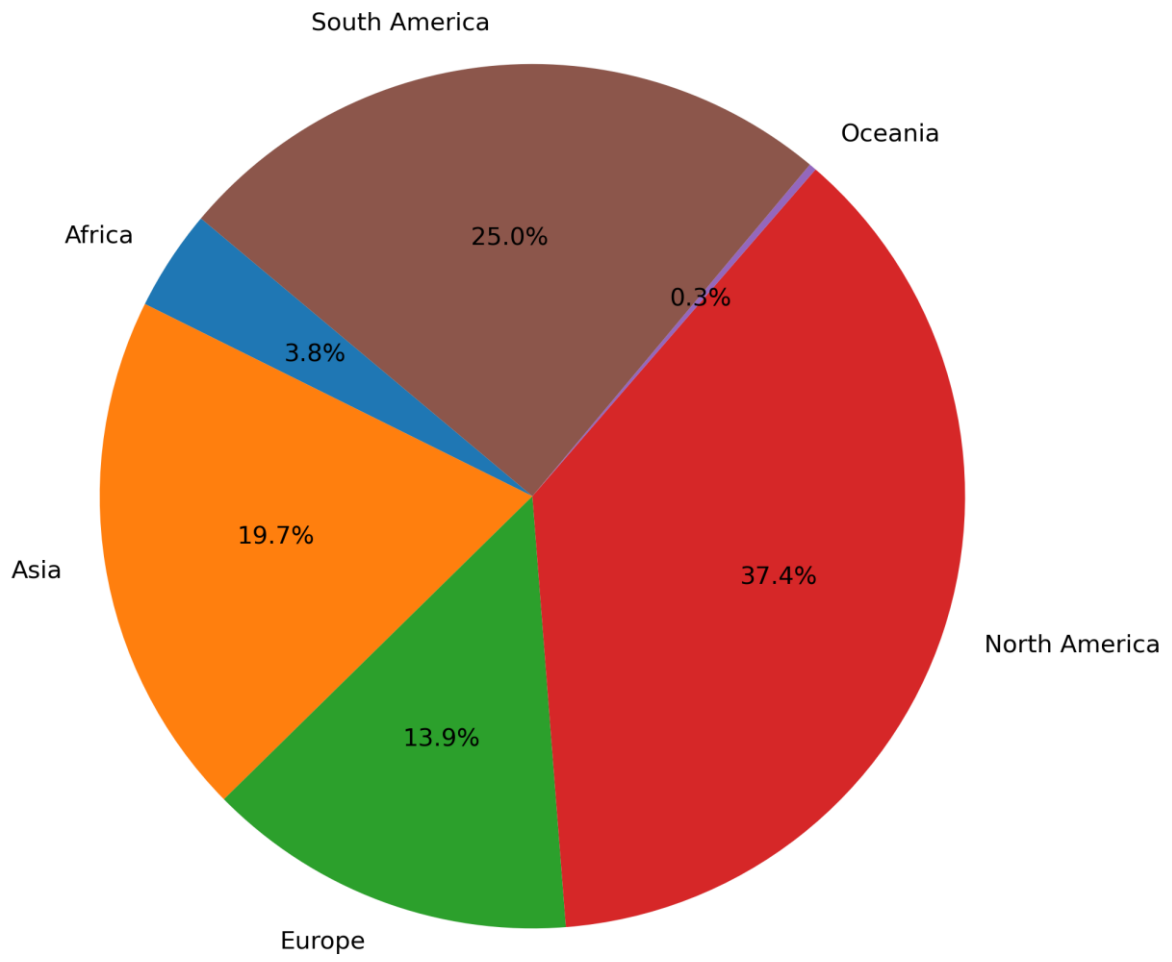
Detaylı bilgi edinme yolunda pie chart kimi zaman yetersiz kalabilmektedir. Bu detaylar veri setinin çarpıklığı, dağılımı, ortalaması, standart sapması ve benzeri bilgiler olabilir. Ayrıyeten eğer veri setinde bulunan sınıfların sayısı fazla ise görsel olarak ayırt etmesi zor ve karşılaştırması güç görseller çıkartabilir. Tüm bu durumlar kendisine spesifik bir görevde kullanım alanı sağlamıştır.

Pie Chart'ın Kullanım Alanları

Pie chart'ın kullanım şekli oldukça basit olduğundan dolayı kullanım alanı da bir o kadar geniş olmaktadır. **Şekil 10**'da da görüldüğü üzere bir veri setinin sınıfları özelinde karşılaştırmalar yapılmıştır. Bu karşılaştırma kıtalarda gerçekleşen toplam ölümleri ele almaktadır. Görüldüğü üzere verilerin dağılımına dair herhangi bir bilgi sağlanmasa da oransal olarak karşılaştırma özelinde güzel sonuçlar elde edilmiştir. Buradaki dağılımdan kasıt ölümlerin her bir gün bazında gösterilmeyişidir.

- Kategorik Verilerin Dağılımı
- Sosyal ve Demografik Araştırmalar
- Anket Sonuçları
- Bilimsel Araştırmalar
- Vb.

Alanlarda pasta grafiğinin kullanım alanı oldukça yaygındır ve kullanılmaya devam edilmektedir.



Şekil 10