**Introduction**

**Introduction to Statistics**

Every day, we encounter numerical information that describes or analyzes the world. For example, here are some news items that were published after the coronavirus appeared in China.

"Among more than 44,000 confirmed cases, there were few detected infections among children aged 9 and younger — only 416 or about 1% of the total cases. None of them died.

Even cases among children and teens aged 9 to 19 were rare; there were 549 cases in that group, representing 1.2% of the entire study group. There was a single death in that age group. By contrast, 20% of the roughly 1,400 people 80 and older who contracted the disease died."

As new technologies enable people to process and interpret ever-increasing amounts of data, statistics is playing an increasingly important part of many processes of **decision making** today.

In simple terms, statistics is the grammar of science. Here are other definitions of statistics:

According to Wikipedia:

Statistics is the discipline that concerns the collection, organization, analysis, interpretation and presentation of data.

İstatistik, verilerin toplanması, düzenlenmesi, analizi, yorumlanması ve sunulması ile ilgili disiplindir.

According to Merriam-Webster's Collegiate Dictionary:

Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.

İstatistik, sayısal veri yığınlarının toplanması, analizi, yorumlanması ve sunumu ile ilgilenen bir matematik dalıdır.



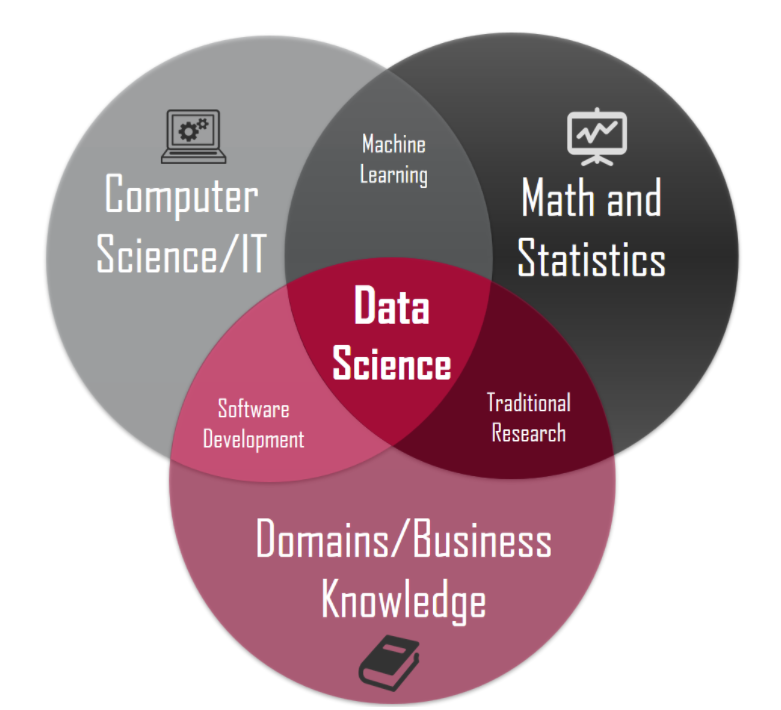
**Why Should You Learn Statistics**

The American Statistical Association defines the statistics as “the science of learning from data and of measuring, controlling, and communicating uncertainty". The relationship between data and statistics is very strong because data in the real world can be processed by statistics and meaningful trends can be obtained from it by performing mathematical calculations within the framework of statistics.  
  
The definition of "a person who knows more statistics than a programmer and more programming than a statistician" is one of the famous descriptions for a data scientist. On the other hand, data science is at the intersection of computer science, math&statistical skills, and domain knowledge. The following picture illustrates the relationship between statistics and data science.

**Neden İstatistik Öğrenmelisiniz?**

Amerikan İstatistik Derneği, istatistikleri "verilerden öğrenme ve belirsizliği ölçme, kontrol etme ve iletme bilimi" olarak tanımlar. Veri ve istatistik arasındaki ilişki çok güçlüdür, çünkü gerçek dünyadaki veriler istatistikler tarafından işlenebilir ve anlamlı eğilimler olabilir. istatistik çerçevesinde matematiksel hesaplamalar yapılarak ondan elde edilebilir.

"Bir programcıdan daha fazla istatistik, bir istatistikçiden daha fazla programlama bilen kişi" tanımı, bir veri bilimci için ünlü tanımlardan biridir. Öte yandan, veri bilimi bilgisayar bilimi, matematik ve istatistik becerileri ve alan bilgisinin kesiştiği noktadadır. Aşağıdaki resim istatistik ve veri bilimi arasındaki ilişkiyi göstermektedir.



Therefore, statistics have an important place in the data science world. We will not discuss statistical topics in-depth like a statistician in this course. However, we will cover the basics of statistics as much as a scientist should know.

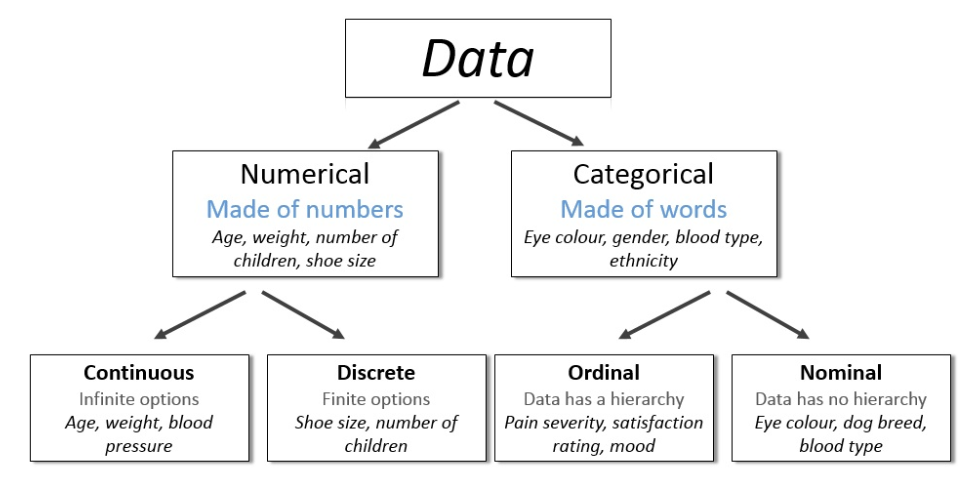
Bu nedenle istatistik, veri bilimi dünyasında önemli bir yere sahiptir. Bu derste istatistik konularını bir istatistikçi gibi derinlemesine tartışmayacağız. Ancak biz bir bilim adamının bilmesi gerektiği kadar istatistiğin temellerini ele alacağız.



## Types of Data

Scientists seek to answer questions using rigorous methods and careful observations. These observations - collected from the likes of field notes, surveys, and experiments - form the backbone of a statistical investigation and are called **data**. Statistics is the study of how best to collect, analyze, and draw conclusions from data.

Bilim adamları, titiz yöntemler ve dikkatli gözlemler kullanarak soruları yanıtlamaya çalışırlar. Alan notları, anketler ve deneyler gibi toplanan bu gözlemler, istatistiksel bir araştırmanın omurgasını oluşturur ve veri olarak adlandırılır. İstatistik, verilerin en iyi nasıl toplanacağını, analiz edileceğini ve bunlardan nasıl sonuçlar çıkarılacağının incelenmesidir.



**📝Data**  
Data are characteristics or information, usually numerical, that are collected through observation.

Veriler, gözlem yoluyla toplanan, genellikle sayısal olan özellikler veya bilgilerdir.



Understanding different types of data is a crucial prerequisite for doing **Exploratory Data Analysis (EDA)** because you can only use certain statistical measurements for certain types of data.

Farklı veri türlerini anlamak, Keşifsel Veri Analizi (EDA) yapmak için çok önemli bir ön koşuldur, çünkü belirli veri türleri için yalnızca belirli istatistiksel ölçümleri kullanabilirsiniz.



There are two main types of data types: **numerical**(or quantitative data) and **categorical**(or qualitative) data.

## Types of Data

### Numerical Data (Quantative)

Numerical data includes **counts** or **measurements**. It can be further classified into **discrete** and **continuous** data.

**Discrete** data contains only finite values. For example the number of students in a school, the number of buildings in a city, the number of desks in a classroom, etc. A possible number of students in a school cannot be infinite. It can be 300, 301 or 302, but it cannot be something between 300 and 301. You can somehow count the values it can take.

**Continuous** data can have an infinite number of values between any two values. For example age, height or weight of a person, the time assigned to a special task, house price, stock price, etc. A stock price might be 10, 11 or 10.4565 dollars, so it can have an infinite number of values. The age of a person is a continuous data because your exact age might be 28, 28.1 or 28.5.

Sayısal veriler, sayıları veya ölçümleri içerir. Ayrık ve sürekli veriler olarak da sınıflandırılabilir.

Ayrık veriler yalnızca sonlu değerleri içerir. Örneğin bir okuldaki öğrenci sayısı, bir şehirdeki bina sayısı, bir sınıftaki sıra sayısı vb. Bir okuldaki olası öğrenci sayısı sonsuz olamaz. 300, 301 veya 302 olabilir ama 300 ile 301 arasında olamaz. Alabileceği değerleri bir şekilde sayabilirsiniz.

Sürekli veriler, herhangi iki değer arasında sonsuz sayıda değere sahip olabilir. Örneğin, bir kişinin yaşı, boyu veya ağırlığı, özel bir göreve ayrılan süre, ev fiyatı, hisse senedi fiyatı vb. Bir hisse senedi fiyatı 10, 11 veya 10.4565 dolar olabilir, dolayısıyla sonsuz sayıda değere sahip olabilir. Bir kişinin yaşı sürekli bir veridir çünkü tam yaşınız 28, 28,1 veya 28,5 olabilir.

**Discrete and Continuous Data**

* A numerical variable is discrete if its possible values form a set of separate numbers, such as 0, 1, 2, 3, ...
* A numerical variable is continuous if its possible values form an interval.

Ayrık ve Sürekli Veri

* Sayısal değişken, olası değerleri 0, 1, 2, 3, ... gibi bir dizi ayrı sayı oluşturuyorsa, **discrete** tir.
* Sayısal bir değişken, olası değerleri bir aralık oluşturuyorsa **continuous** tur

## Types of Data

### Categorical Data

Categorical data can only take a specific set of values representing a set of possible categories, such as female/male genders, automatic/semi-automatic or manual gearboxes.

You can assign numbers to categorical data, for example, you can use 0 for male and 1 for female. But even in this case, numbers do not have actual numerical meaning. They represent the categories.

Categorical data can be further classified into **nominal** and **ordinal** data. Nominal data simply names something without an order being given. Female/male genders, automatic/semi-automatic or manual gearboxes are examples for nominal data.

Unlike nominal data, ordinal data requires an order. For example, suppose you had dinner in a restaurant and they asked if you could participate in the survey. You might rank the taste of the meal as "1" for bad, "2" for average and "3" for good. The survey data is a good example of ordinal data. Here we use numbers for categorical data but as we mentioned above, they do not have actual numerical meaning. Another example can be given about body sizes for a shirt. Small, Medium, Large sizes all represent the ordinal data. Because there is an order between these categories.

Kategorik veriler, yalnızca kadın/erkek cinsiyetleri, otomatik/yarı otomatik veya düz vites kutuları gibi bir dizi olası kategoriyi temsil eden belirli bir değer kümesi alabilir.

Kategorik verilere numara atayabilirsiniz, örneğin erkek için 0, kadın için 1 kullanabilirsiniz. Ancak bu durumda bile sayıların gerçek sayısal anlamı yoktur. Kategorileri temsil ederler.

Kategorik veriler ayrıca nominal ve ordinal (sıralı) olarak sınıflandırılabilir. Nominal veriler, bir sıra verilmeden basitçe bir şeyi adlandırır. Kadın/erkek cinsiyet, otomatik/yarı otomatik veya düz vites kutuları nominal verilere örnektir.

Nominal verilerden farklı olarak, ordinal veriler bir sıra gerektirir. Örneğin, bir restoranda akşam yemeği yediğinizi ve ankete katılıp katılamayacağınızı sorduklarını varsayalım. Yemeğin lezzetini "1" kötü, "2" ortalama ve "3" iyi olarak sıralayabilirsiniz. Anket verileri, sıralı verilere iyi bir örnektir. Burada kategorik veriler için sayılar kullanıyoruz ancak yukarıda bahsettiğimiz gibi gerçek sayısal anlamları yoktur. Bir gömlek için beden ölçüleri hakkında başka bir örnek verilebilir. Küçük, Orta, Büyük boyutların tümü sıralı verileri temsil eder. Çünkü bu kategoriler arasında bir düzen vardır.

**📝Ordinal and Nominal Data**

* An ordinal data is a categorical one for which the categories are ordered from low to high in some sense.
* A nominal data is a categorical data that do not have a natural order or ranking.

• Ordinal (Sıralı veri), kategorilerin bir anlamda düşükten yükseğe doğru sıralandığı kategorik bir veridir.

• Nominal veri, doğal bir sıralaması veya sıralaması olmayan kategorik bir veridir.

## Level of Measurements

### Description

The level of measurement or scale of measure is a classification describing the essence of the information within the values assigned to variables.

Ölçme düzeyi veya ölçü ölçeği, değişkenlere atanan değerler içindeki bilgilerin özünü açıklayan bir sınıflandırmadır.



**Variable**  
A variable is any characteristic observed in a study.

Değişken, bir çalışmada gözlemlenen herhangi bir özelliktir.

There are four different levels of measurements: **nominal, ordinal, interval, or ratio**. As we will recall from the previous lesson, the nominal and ordinal data are two subcategories of categorical data. Correspondingly we will use one of the nominal or ordinal levels of measurements for categorical data, and interval or ratio level of measurements for numerical data. Here the important point is that there is a hierarchy between these levels, and this hierarchy increases from nominal to ratio level.

A data scientist needs to understand the different levels of measurement because that dictates what statistical analysis is appropriate for our data.

Dört farklı ölçüm seviyesi vardır: nominal, sıra, aralık veya oran. Bir önceki dersten hatırlayacağımız gibi, nominal ve ordinal veriler, kategorik verilerin iki alt kategorisidir. Buna uygun olarak, kategorik veriler için nominal veya sıralı ölçüm düzeylerinden birini ve sayısal veriler için aralık veya oran ölçüm düzeylerini kullanacağız. **Burada önemli olan nokta, bu seviyeler arasında bir hiyerarşi olması ve bu hiyerarşinin nominalden oran seviyesine doğru artmasıdır.**

Bir veri bilimcisinin farklı ölçüm düzeylerini anlaması gerekir çünkü bu, verilerimiz için hangi istatistiksel analizin uygun olduğunu belirler.

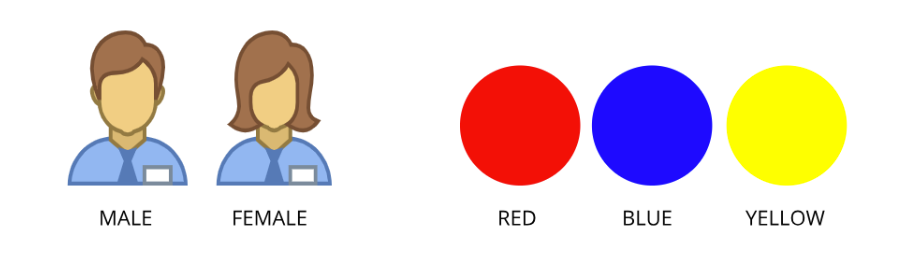
## Level of Measurements

### Nominal Level of Measurement

The first level is the nominal level of measurement. As we mentioned earlier, the nominal level simply names something without assigning an order. In this level, the numbers in the variable are used only to classify the data. For example, using 0 for male and 1 for female. Male/Female, Red/Blue/Yellow or Success/Fail classifications are examples for the nominal level of measurement.

İlk seviye, nominal ölçüm seviyesidir. Daha önce de belirttiğimiz gibi, nominal düzey, bir sıra atamadan basitçe bir şeyi adlandırır. Bu seviyede, **değişkendeki sayılar sadece verileri sınıflandırmak için kullanılır**. Örneğin, erkek için 0 ve kadın için 1 kullanılması. Erkek/Kadın, Kırmızı/Mavi/Sarı veya Başarılı/Başarısız sınıflandırmaları, nominal ölçüm düzeyi için örneklerdir.





**Tips:**

* Nominal Level of Measurement simply names something without assigning an order.
* • Nominal Ölçüm Düzeyi, bir sıra atamadan basitçe bir şeyi adlandırır.
* 

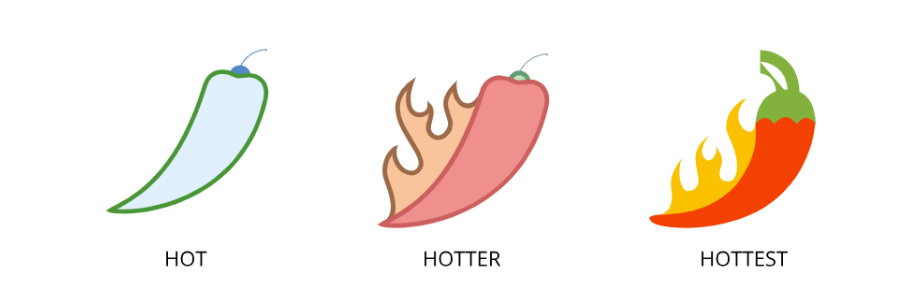
## Level of Measurements

### Ordinal Level of Measurement

The second level is the ordinal level of measurement. In this level, the attributes are ordered, however, distances between attributes have no meaning. Remember the survey in the restaurant. In this measure, higher numbers mean more taste. But the distance from 1 to 2 might not be the same as 2 to 3. In the ordinal level of measurement, the interval between values is not interpretable.

İkinci seviye, sıralı ölçüm seviyesidir. **Bu seviyede, nitelikler sıralanır, ancak nitelikler arasındaki mesafelerin bir anlamı yoktur**. Restorandaki anketi hatırlayın. Bu ölçümde, daha yüksek sayılar daha fazla tat anlamına gelir. Ancak 1'den 2'ye olan mesafe 2'den 3'e kadar aynı olmayabilir. Sıralı ölçüm düzeyinde, değerler arasındaki aralık yorumlanamaz.





**Tips:**

* In the ordinal level, the attributes are ordered.
* But distances between attributes have no meaning.
* **Sıra düzeyinde, nitelikler sıralanır.**
* **Ancak nitelikler arasındaki mesafelerin bir anlamı yoktur.**

**Level of Measurements**

**Interval Level of Measurement**

Unlike the ordinal scale, in the interval level of measurement, the distance between attributes does have meaning. For example, while measuring temperature, the distance between 40-50 degrees is the same as the distance between 60-70 degrees. The interval between values can be interpreted. That means, it makes sense to calculate an average of an interval, where it doesn't work for ordinal scales. However, in the interval level of measurement the **ratio** does not make any sense. It is not possible to say that 100 degrees are twice as hot as 50 degrees.

**Ölçüm Düzeyi**

**Aralık Ölçüm Seviyesi**

Sıralı ölçeğin aksine, ölçüm aralığı düzeyinde, **nitelikler arasındaki mesafenin bir anlamı vardır**. Örneğin sıcaklık ölçerken 40-50 derece arasındaki mesafe ile 60-70 derece arasındaki mesafe aynıdır. **Değerler arasındaki aralık yorumlanabilir**. Bu, sıralı ölçekler için çalışmadığı bir aralığın ortalamasını hesaplamanın mantıklı olduğu anlamına gelir. Ancak, ölçüm aralığı düzeyinde **oran bir anlam ifade etmemektedir**. 100 derecenin 50 derecenin iki katı olduğunu söylemek mümkün değildir.

**💡Tips:**

* The distance between attributes does have meaning.
* But the ratio does not make any sense.

Nitelikler arasındaki mesafenin anlamı vardır.

Ama oran bir anlam ifade etmiyor.

**Level of Measurements**

**Ratio Level of Measurement**

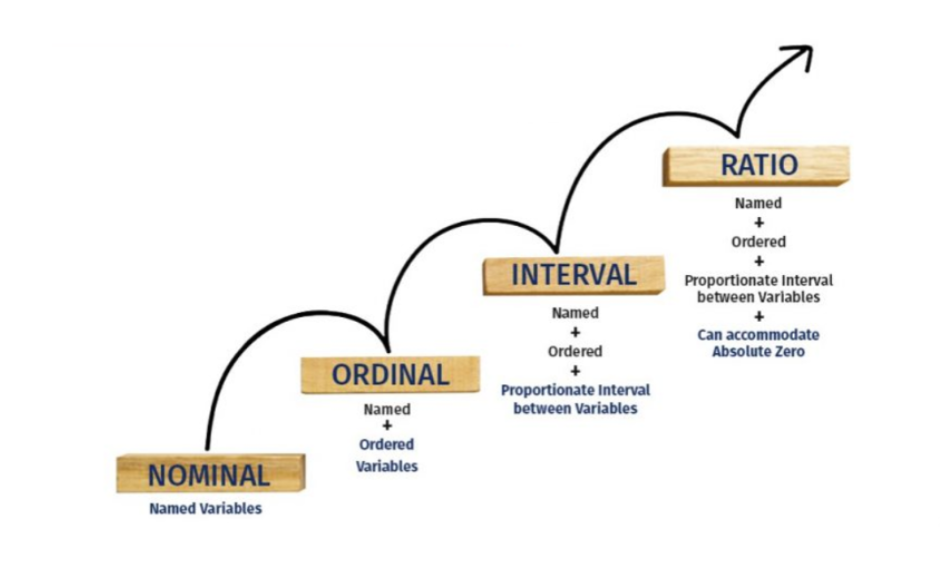
The fourth level is the ratio level of measurement. In this level, the observations, besides having the same intervals, also have a zero value. The zero value makes this level of measurement different from the interval scale. Examples for this level of measurement include length, height, weight or cell phone charge capacity. Unlike the interval scales, ratios are meaningful. It is possible to say that one object has twice the height. Because having a zero-point makes it meaningful.

Dördüncü seviye, Ratio (ölçüm oranı) seviyesidir. Bu düzeyde, **gözlemler aynı aralıklara sahip olmanın yanı sıra sıfır değerine de sahiptir**. Sıfır değeri, bu ölçüm düzeyini aralık ölçeğinden farklı kılar. Bu ölçüm düzeyine örnek olarak uzunluk, boy, ağırlık veya cep telefonu şarj kapasitesi verilebilir. **Aralıklı ölçeklerden farklı olarak oranlar anlamlıdır**. Bir cismin yüksekliğinin iki katı olduğunu söylemek mümkündür. Çünkü sıfır noktasına sahip olmak onu anlamlı kılar.

**💡Tips:**

* Ratios make sense.
* There is a zero point.
* Oranlar mantıklı.
* Sıfır noktası var.

The following picture shows the level of measurements and the hierarchy between them.



## Graphical Representation of Data

### Data Patterns in Statistics

Graphical representations are useful for viewing patterns in data. In general, we can define the patterns in data as follows: center, spread, shape, and unusual features.  
Some known distributions have some special descriptive labels such as symmetric, bell-shaped, skewed, etc.  
  
Grafik temsiller, verilerdeki kalıpları görüntülemek için kullanışlıdır. Genel olarak verilerdeki örüntüleri şu şekilde tanımlayabiliriz: merkez, yayılma, şekil ve sıra dışı özellikler.

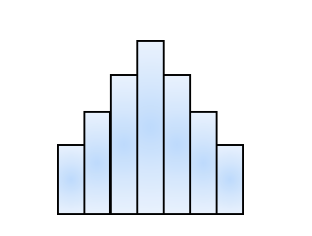
Bilinen bazı dağılımlar, simetrik, çan şeklinde, çarpık vb. gibi bazı özel tanımlayıcı etiketlere sahiptir.

#### **Center**

The center of a distribution, graphically, is located at the median of the distribution. Such a graphic chart displays that almost half of the observations are on either side. Height of each column indicates the frequency of observations.

Bir dağılımın merkezi, grafiksel olarak dağılımın medyanında bulunur. Böyle bir grafik çizelge, gözlemlerin neredeyse yarısının her iki tarafta olduğunu gösterir. Her sütunun yüksekliği, gözlemlerin sıklığını gösterir.



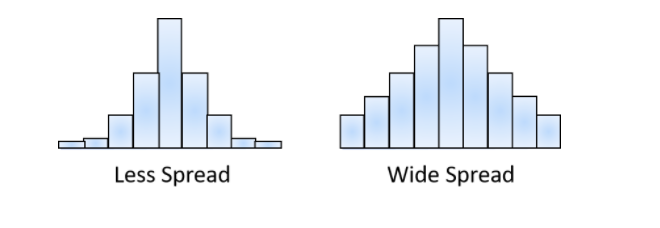


#### **Spread**

If we talk about the spread of a distribution, the variation of the data will be understood. If the set of observations spreads over a wide range, the spread is greater. If the observations are centered around a single value in a narrower range, then the spread is smaller.

Bir dağılımın yayılmasından bahsedersek, verilerin varyasyonu anlaşılır. Gözlem kümesi geniş bir aralığa yayılıyorsa, yayılma daha büyüktür. Gözlemler daha dar bir aralıkta tek bir değer etrafında ortalanırsa, yayılma daha küçüktür.



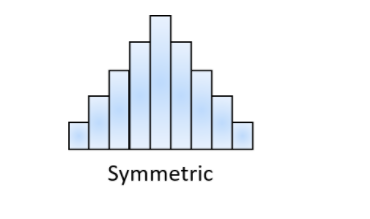


#### **Shape**

The shape of a distribution can described using following characteristics.

Bir dağılımın şekli aşağıdaki özellikler kullanılarak tanımlanabilir:

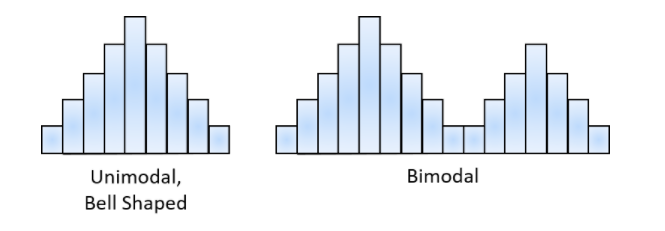
* **Symmetry** - In symmetric distribution, graph can be divided at the center in such a way that each half is a mirror image of the other. In the Normal Distribution, the two sides of the distribution are equal and symmetrical.
* **Simetri** - Simetrik dağılımda, grafik, her yarım diğerinin ayna görüntüsü olacak şekilde merkezde bölünebilir. Normal Dağılımda dağılımın iki tarafı eşit ve simetriktir.



**Number of peaks.** - Distributions with one or multiple peaks. Distribution with one clear peak is known as unimodal, and **distribution with two clear peaks is called bimodal.** **A single peak symmetric distribution at the center, is referred to as bell-shaped.**

**Pik sayısı**. - Bir veya birden fazla tepe noktası olan dağılımlar. Bir net tepe ile dağıtım tek modlu olarak bilinir ve iki açık tepe ile dağıtım iki modlu olarak bilinir. Merkezde tek bir tepe simetrik dağılım, çan şeklinde olarak adlandırılır.





**Skewness** - Some distributions may have multiple observations on one side of the graph than the other side. Distributions having fewer observations towards higher values are said to be skewed right; and distributions with fewer observations towards lower values are said to be skewed left.

**Çarpıklık** - Bazı dağılımlar, grafiğin bir tarafında diğer tarafında olduğundan daha fazla gözleme sahip olabilir. Daha yüksek değerlere yönelik daha az gözlemi olan dağılımların sağa çarpık olduğu söylenir; ve daha düşük değerlere doğru daha az gözlemi olan dağılımların sola çarpık olduğu söylenir.

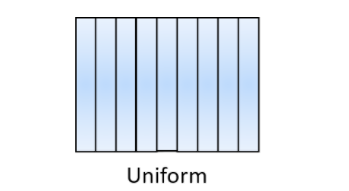




**Uniform** - When the set of observations has no peak and have data equally spread across the range of the distribution, then the distribution is called a uniform distribution.

**Tekdüze** - Gözlem kümesinin tepe noktası olmadığında ve verilerin dağılım aralığına eşit olarak dağıldığı durumlarda, dağılıma tekdüze dağılım denir.



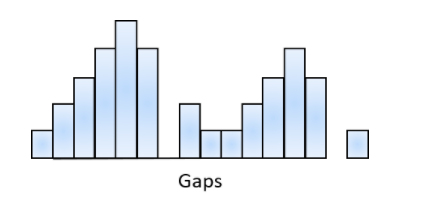


#### **Unusual Features**

* Common unusual features of data patterns are **gaps** and **outliers**.

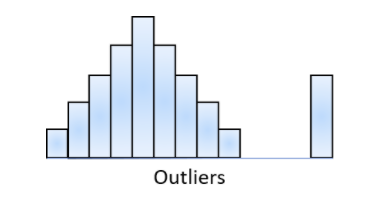
**Olağandışı Özellikler**

* Veri modellerinin ortak olağandışı özellikleri, boşluklar ve aykırı değerlerdir.
* **Gaps** - Gaps points to areas of a distribution having no observations. Following figure has a gap as there are no observations in the middle of the distribution.
* **Boşluklar** - Boşluklar, bir dağılımın gözlemi olmayan alanlarına işaret eder. Dağılımın ortasında gözlem olmadığı için aşağıdaki şekilde bir boşluk vardır.



**Outliers** - Distributions may be characterized by extreme values that differ greatly from the other set of observation data. These extreme values are refered as outliers. Following figure illustrates a distribution with an outlier.

**Aykırı Değerler** - Dağılımlar, diğer gözlem veri setinden büyük ölçüde farklılık gösteren uç değerlerle karakterize edilebilir. Bu uç değerler aykırı değerler olarak adlandırılır. Aşağıdaki şekil bir aykırı değer içeren bir dağılımı göstermektedir.

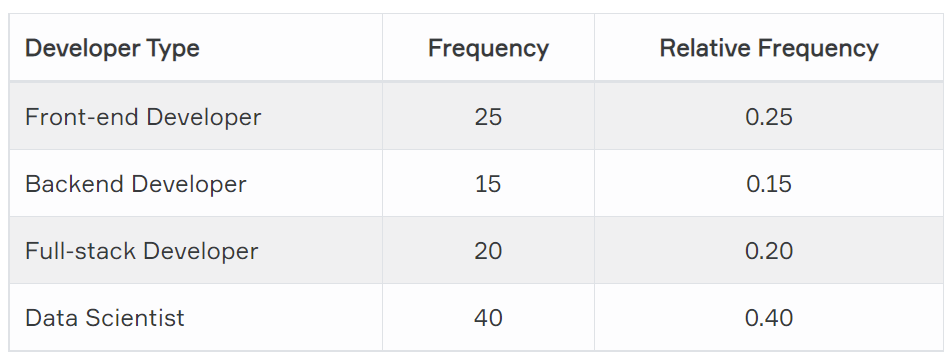


## Graphical Representation of Data

### Frequency Table

A **frequency table** is a table that lists items and shows the number of times the items occur. **The relative frequency** of a category or a numerical value **is the associated frequency divided by the total number of data**. The table below shows the frequencies of the various response categories. It also shows the relative frequencies, which are the proportion of responses in each category. For example, the relative frequency for "Front-end Developer" is 25/100 = 0.25.

**Frequency Table** öğeleri listeleyen ve öğelerin kaç kez meydana geldiğini gösteren bir tablodur. Bir kategorinin veya sayısal bir değerin göreli frekansı, toplam veri sayısına bölünen ilişkili frekanstır. Aşağıdaki tablo, çeşitli yanıt kategorilerinin frekanslarını göstermektedir. Ayrıca, her kategorideki yanıtların oranı olan göreli frekansları da gösterir. Örneğin, " Front-end Developer " için göreli frekans 25/100 = 0,25'tir.



**Frequency Table**  
A frequency table is a listing of possible values for a variable, together with the number of observations for each value.

**Frekans tablosu**

Sıklık tablosu, her bir değer için gözlem sayısıyla birlikte bir değişken için olası değerlerin bir listesidir.

## Graphical Representation of Data

### Pie Chart

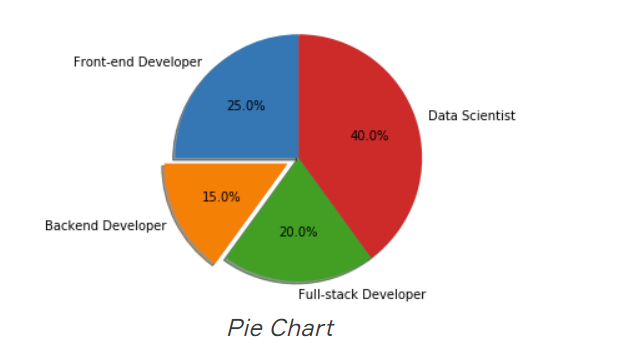
In a pie chart, each category is represented by a slice of the pie. The area of the slice is proportional to the percentage of responses in the category. This is simply the relative frequency multiplied by 100.

Pasta grafiğinde, her kategori pastanın bir dilimi ile temsil edilir. Dilimin alanı, kategorideki yanıtların yüzdesiyle orantılıdır. Bu, 100 ile çarpılan nispi frekanstır.

**📝Pie Chart**  
A pie chart is a circle having a “slice of the pie” for each category.

Pasta grafiği, her kategori için "pastadan bir dilim" içeren bir dairedir

As can be seen from the pie chart below, most of the developers in the company are Data Scientists (40%) , 25% of the developers are Front-end Developers, 20% Full-stack Developers, and 15% Backend Developers.



Pie charts are effective for displaying the relative frequencies of a small number of categories. They are not recommended, however, when you have a large number of categories. Pie charts can also be confusing when they are used to compare the outcomes of two different surveys or experiments.

Pasta grafikler, az sayıda kategorinin göreli frekanslarını görüntülemek için etkilidir. Bununla birlikte, çok sayıda kategoriniz olduğunda önerilmezler. Pasta grafikler, iki farklı anket veya deneyin sonuçlarını karşılaştırmak için kullanıldıklarında da kafa karıştırıcı olabilir.

**Tips:**

* When slices become too small, pie charts have to rely on colors, textures or arrows so the reader can understand them.
* Dilimler çok küçük olduğunda, pasta grafikler okuyucunun anlayabilmesi için renklere, dokulara veya oklara dayanmalıdır.

**⚠️Avoid ! :**

* Pie charts are not recommended if there are too many categories of data being presented.
* • Sunulan çok fazla veri kategorisi varsa pasta grafikler önerilmez
* 

## Graphical Representation of Data

### Bar Chart

Bar charts can be used to represent frequencies of different categories. A bar chart is made up of columns plotted on a graph.

Çubuk grafikler, farklı kategorilerin frekanslarını temsil etmek için kullanılabilir. Bir çubuk grafik, bir grafik üzerine çizilen sütunlardan oluşur



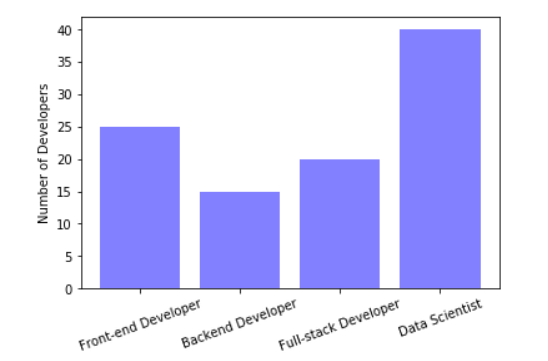
**📝Bar Chart**  
A bar chart is a chart that presents categorical data with rectangular bars with heights proportional to the values that they represent.

Çubuk grafik, temsil ettikleri değerlerle orantılı yüksekliklere sahip dikdörtgen çubuklarla kategorik verileri sunan bir grafiktir.

A bar chart of the developers employed in a company is shown below. Frequencies are shown on the Y-axis and the type of developers is shown on the X-axis. Typically, the Y-axis shows the number of observations in each category rather than the percentage of observations as is typical in pie charts.

Bir şirkette çalışan geliştiricilerin bir çubuk grafiği aşağıda gösterilmiştir. Frekanslar Y ekseninde gösterilir ve geliştiricilerin türü X ekseninde gösterilir. Tipik olarak, Y ekseni, pasta grafiklerde olduğu gibi gözlemlerin yüzdesinden ziyade her kategorideki gözlemlerin sayısını gösterir.





Here is how to read a bar chart.

* The columns are positioned over a label that represents a categorical variable.
* The height of the column indicates the size of the group defined by the column label.

Bir çubuk grafiği nasıl okuyacağınız aşağıda açıklanmıştır.

Sütunlar, kategorik bir değişkeni temsil eden bir etiketin üzerine yerleştirilir.

Sütunun yüksekliği, sütun etiketi tarafından tanımlanan grubun boyutunu gösterir.

**Tips:**

* The bars can be plotted vertically or horizontally.
* A vertical bar chart is sometimes called a column chart.

**⚠️Avoid ! :**

* Do not start the axis with a value above zero. If you do this, you shrink the bars and get confusing visuals.
* Do not use the rainbow effect. Although coloring the graphics sounds good, it often makes the tables difficult to understand.

İpuçları:

Çubuklar dikey veya yatay olarak çizilebilir.

Dikey çubuk grafiğe bazen sütun grafiği denir.

⚠️ Kaçının ! :

Ekseni sıfırın üzerinde bir değerle başlatmayın. Bunu yaparsanız çubukları küçültür ve kafa karıştırıcı görseller elde edersiniz.

Gökkuşağı efektini kullanmayın. Grafikleri renklendirmek kulağa hoş gelse de, genellikle tabloların anlaşılmasını zorlaştırır.

## Graphical Representation of Data

### Histogram

A histogram is a graphical method for displaying the shape of a distribution. It is particularly useful when there are a large number of observations. We begin with an example consisting of the scores of 250 students on a statistics test. The students' scores ranged from 32 to 96.

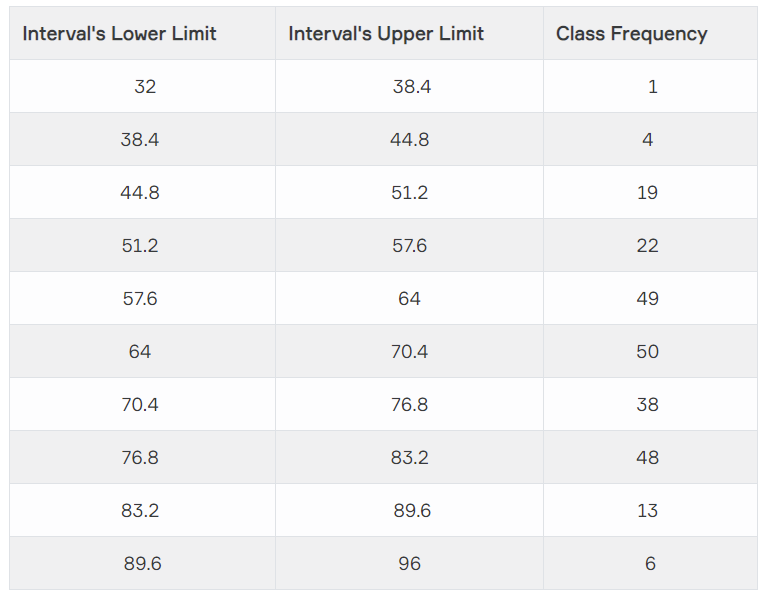
Histogram, bir dağılımın şeklini görüntülemek için grafiksel bir yöntemdir. Çok sayıda gözlem olduğunda özellikle yararlıdır. Bir istatistik testinde 250 öğrencinin puanlarından oluşan bir örnekle başlıyoruz. Öğrencilerin puanları 32 ile 96 arasında değişiyordu.

**📝Histogram**  
A histogram is a graphical representation of a distribution.

The first step is to create a frequency table. Unfortunately, a simple frequency table would be too big, containing over 100 rows. To simplify the table, we group scores together as shown in table below.

İlk adım, bir sıklık tablosu oluşturmaktır. Ne yazık ki, 100'den fazla satır içeren basit bir sıklık tablosu çok büyük olacaktır. Tabloyu basitleştirmek için puanları aşağıdaki tabloda gösterildiği gibi gruplandırıyoruz.





To create this table, the range of scores was broken into intervals, called class intervals. The first interval is from 32 to 38.4, the second from 38.4 to 44.8, etc. Next, the number of scores falling into each interval was counted to obtain the class frequencies. There are one score in the first interval, 4 in the second, etc. The widths of the class intervals, sometimes called bin widths. Your choice of bin width determines the number of class intervals.

In a histogram, the class frequencies are represented by bars. The height of each bar corresponds to its class frequency. A histogram of these data is shown below.

Bu tabloyu oluşturmak için puan aralığı, sınıf aralıkları adı verilen aralıklara bölündü. İlk aralık 32'den 38.4'e, ikincisi 38.4'ten 44.8'e, vb. Daha sonra, sınıf frekanslarını elde etmek için her aralığa düşen puanların sayısı sayıldı. İlk aralıkta bir puan, ikincide 4 puan vb. vardır. Sınıf aralıklarının genişlikleri, bazen bin genişlikleri olarak adlandırılır. Bölme genişliği seçiminiz, sınıf aralıklarının sayısını belirler.

Bir histogramda, sınıf frekansları çubuklarla temsil edilir. Her çubuğun yüksekliği, sınıf frekansına karşılık gelir. Bu verilerin bir histogramı aşağıda gösterilmiştir.



The histogram makes it plain that most of the scores are in the middle of the distribution, with fewer scores in the extremes. You can see from the chart that frequency is greatest in the 64 to 70.4 score group.

Histogram, puanların çoğunun, aşırı uçlarda daha az puanla, dağılımın ortasında olduğunu açıkça ortaya koymaktadır. Tablodan, frekansın 64 ila 70.4 puan grubunda en yüksek olduğunu görebilirsiniz.

**💡Tips:**

* With bar charts, the labels on the X axis are categorical; with histograms, the labels are quantitative.

• Çubuk grafiklerde X eksenindeki etiketler kategoriktir; histogramlarla, etiketler niceldir.



## Population & Sample

### Population

Population means the aggregation of all the elements under study with one or more common features. The population is not confined to people only, also animals, objects, buildings, cars, etc can constitute the population.

Popülasyon, incelenen tüm unsurların bir veya daha fazla ortak özellik ile toplanması anlamına gelir. Nüfus sadece insanlarla sınırlı değildir, ayrıca hayvanlar, nesneler, binalar, arabalar vb. popülasyonu oluşturabilir.



**📝Population**  
A population is the total group about whom you want to make conclusions.

population, hakkında sonuç çıkarmak istediğiniz toplam gruptur.



For example, all people living in the USA, all buildings in a country, disabled children in India, all students in the world, diesel cars in Europe, all sheep in the USA can constitute a population. It can be of any size, and the number of elements or members in a population is known as population size. If there are a hundred million people in a country, then the population size is 100 million.

Örneğin ABD'de yaşayan tüm insanlar, bir ülkedeki tüm binalar, Hindistan'daki engelli çocuklar, dünyadaki tüm öğrenciler, Avrupa'daki dizel arabalar, ABD'deki tüm koyunlar bir popülasyon oluşturabilir. Herhangi bir boyutta olabilir ve bir popülasyondaki eleman veya üye sayısı popülasyon büyüklüğü olarak bilinir. Bir ülkede yüz milyon insan varsa, nüfus büyüklüğü 100 milyondur.



## Population & Sample

### Sample

While conducting a statistical study, samples are mainly used when the population size is too large to include all the members of the population. For example, it is easier to conduct a study with 2000 people selected from 100 million people, rather than conducting a study involving 100 million people. The selected respondents from the population constitute a sample and the selection process is known as sampling. The elements under study are called sampling units, and the number of units in a sample is called a sample size.

İstatistiksel bir çalışma yürütülürken, numuneler esas olarak popülasyon büyüklüğü popülasyonun tüm üyelerini içeremeyecek kadar büyük olduğunda kullanılır. Örneğin 100 milyon kişiden oluşan bir araştırma yapmaktansa, 100 milyon kişiden seçilen 2000 kişi ile araştırma yapmak daha kolaydır. Popülasyondan seçilen katılımcılar bir örneklem oluşturur ve seçim süreci örnekleme olarak bilinir. İncelenen öğelere örnekleme birimleri denir ve bir örnekteki birim sayısına örnek boyutu denir.



**Sample**  
A sample is a subset of the population for whom you actually have data

Örnek, gerçekten veriye sahip olduğunuz popülasyonun bir alt kümesidir.

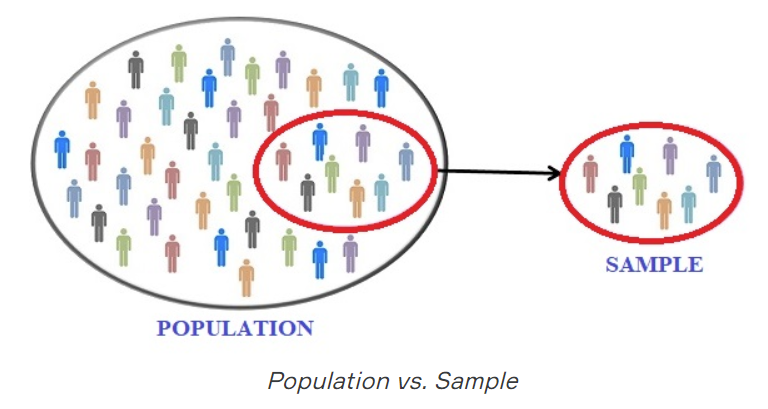


To sum up; the population represents the whole members with certain features. On the other hand, the sample is a limited subset of the population, selected by a special process. Population is always the target of research. We learn about the population by taking the sample from the collection

The following picture explains the relationship between population and sample very well.

Özetle; popülasyon, belirli özelliklere sahip tüm üyeleri temsil eder. Öte yandan, örneklem, özel bir işlemle seçilen popülasyonun sınırlı bir alt kümesidir. popülasyon her zaman araştırmanın hedefidir. Koleksiyondan örnek alarak popülasyon hakkında bilgi ediniriz.

Aşağıdaki resim popülasyon ve örneklem arasındaki ilişkiyi çok iyi açıklamaktadır.



**Tips:**

* We observe samples but are interested in populations.

## Population & Sample

### Population Parameters and Sample Statistics

Of the numerical summaries introduced so far, the mean *x̄* and the standard deviation *s* are the most commonly used in practice. We’ll use them frequently in the rest of the text. The formulas that define *x̄* and *s* refer to sample data. They are sample statistics.

We will distinguish between sample statistics and the corresponding parameter values for the population. The population mean is the average of all observations in the population. The population standard deviation describes the variability of the population observations about the population mean. These are usually unknown. Inferential statistical methods help us to make decisions and predictions about the population parameters based on the sample statistics.

Şimdiye kadar tanıtılan sayısal özetlerden, ortalama x̄ ve standart sapma s, pratikte en yaygın olarak kullanılanlardır. Bunları metnin geri kalanında sık sık kullanacağız. x̄ ve s'yi tanımlayan formüller örnek verilere atıfta bulunur. Bunlar örnek istatistiklerdir.

Örnek istatistikler ile popülasyon için karşılık gelen parametre değerleri arasında ayrım yapacağız. Popülasyon ortalaması, popülasyondaki tüm gözlemlerin ortalamasıdır. Popülasyon standart sapması, popülasyon ortalamasıyla ilgili popülasyon gözlemlerinin değişkenliğini tanımlar. Bunlar genellikle bilinmez. Çıkarımsal istatistiksel yöntemler, örnek istatistiklere dayalı olarak popülasyon parametreleri hakkında kararlar ve tahminler yapmamıza yardımcı olur.

**Tips:**

* A **parameter**is a numerical summary of the population, and a **statistic**is a numerical summary of a sample.
* Bir parametre, popülasyonun sayısal bir özetidir ve bir istatistik, bir örneğin sayısal bir özetidir.