**Introductory Statistics for Data Science**

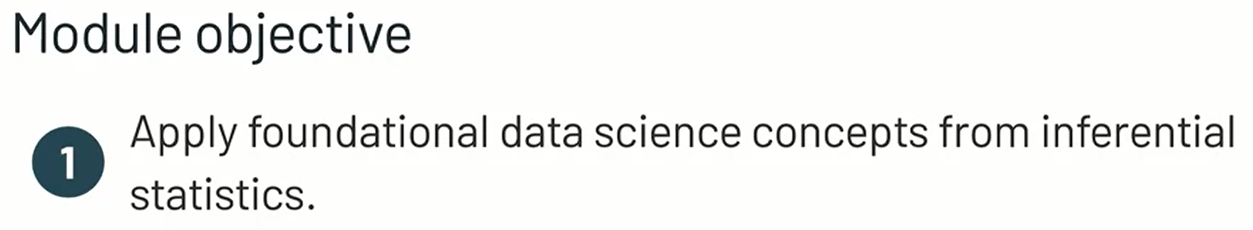
**Learning Objectives**

* Use hypothesis testing to answer a series of questions about the data set [Veri seti hakkında bir dizi soruyu cevaplamak için hipotez testini kullanın]
* Describe how outlier detection techniques can be used to determine if a data point is an outlier given a specific distribution [Bir veri noktasının belirli bir dağılım verilen bir aykırı değer olup olmadığını belirlemek için aykırı değer tespit tekniklerinin nasıl kullanılabileceğini açıklayın]
* Demonstrate the use of hypothesis tests with other hypotheses and other probability distributions [Diğer hipotezler ve diğer olasılık dağılımları ile hipotez testlerinin kullanımını göstermek]
* Demonstrate an understanding of hypothesis testing concepts [Hipotez testi kavramlarını anladığını gösterme]
* Demonstrate the use of a hypothesis test to answer a real-world question using normally-distributed data [Normal olarak dağıtılmış verileri kullanarak gerçek dünyadaki bir soruyu yanıtlamak için bir hipotez testinin kullanımını gösterin]
* Describe hypothesis testing as a tool for connecting inferential statistics, probability and the scientific method [Çıkarımsal istatistik, olasılık ve bilimsel yöntemi birbirine bağlamak için bir araç olarak hipotez testini tanımlayın]
* Identify one real world scenario that could be modeled by a discrete probability distribution and another real world scenario that can be modeled by a continuous probability distribution [Kesintili bir olasılık dağılımı ile modellenebilecek bir gerçek dünya senaryosu ve sürekli bir olasılık dağılımı ile modellenebilecek başka bir gerçek dünya senaryosu tanımlayın.]
* Apply hypothesis testing and outlier techniques to learn from data [Verilerden öğrenmek için hipotez testi ve aykırı değer tekniklerini uygulayın]
* Describe a few common continuous probability distributions and demonstrate how to use them practically to model data [Birkaç yaygın sürekli olasılık dağılımını tanımlayın ve bunların verileri modellemek için pratik olarak nasıl kullanılacağını gösterin]
* Use a Bernoulli distribution to model a real-world scenario [Gerçek dünya senaryosunu modellemek için bir Bernoulli dağılımı kullanın]
* Describe a few common discrete probability distributions [Birkaç yaygın ayrık olasılık dağılımını tanımlayın]
* Demonstrate the practical use of discrete probability distributions to model data [Verileri modellemek için ayrık olasılık dağılımlarının pratik kullanımını gösterin]
* Demonstrate an understanding of probability distributions and how they represent data [Olasılık dağılımlarını ve bunların verileri nasıl temsil ettiğini anladığını gösterme]
* Describe the concept behind probability distributions how they can represent data and random variables [Olasılık dağılımlarının ardındaki kavramı, verileri ve rastgele değişkenleri nasıl temsil edebileceklerini açıklayın]
* Recall foundational concepts about statistics presented in this course [Bu derste sunulan istatistikle ilgili temel kavramları hatırlayın]
* Explain how discrete and continuous probability distributions can be used to model data. [Ayrık ve sürekli olasılık dağılımlarının verileri modellemek için nasıl kullanılabileceğini açıklayın.]
* Compare and contrast discrete and continuous probability distributions [Ayrık ve sürekli olasılık dağılımlarını karşılaştırın ve karşılaştırın]
* Use the basic rules of discrete probability to complete a series of exercises [Bir dizi alıştırmayı tamamlamak için temel ayrık olasılık kurallarını kullanın.]
* Describe the basic rules of discrete probability [Ayrık olasılığın temel kurallarını tanımlayın]
* Compare and contrast descriptive statistics and inferential statistics [Tanımlayıcı istatistikleri ve çıkarımsal istatistikleri karşılaştırın ve karşılaştırın]
* Describe the field of probability and its relationship to inferential statistics [Olasılık alanını ve çıkarımsal istatistiklerle ilişkisini tanımlayın]
* Describe inferential statistics [Çıkarımsal istatistikleri tanımlayın]
* Use descriptive statistics to gather information about a data set [Bir veri seti hakkında bilgi toplamak için tanımlayıcı istatistikleri kullanın]
* Describe descriptive statistics [Tanımlayıcı istatistikleri tanımlayın]
* Describe statistics as a tool for learning from data [İstatistikleri verilerden öğrenmek için bir araç olarak tanımlayın]
* Apply foundational data science concepts from inferential statistics [Çıkarımsal istatistiklerden temel veri bilimi kavramlarını uygulayın]

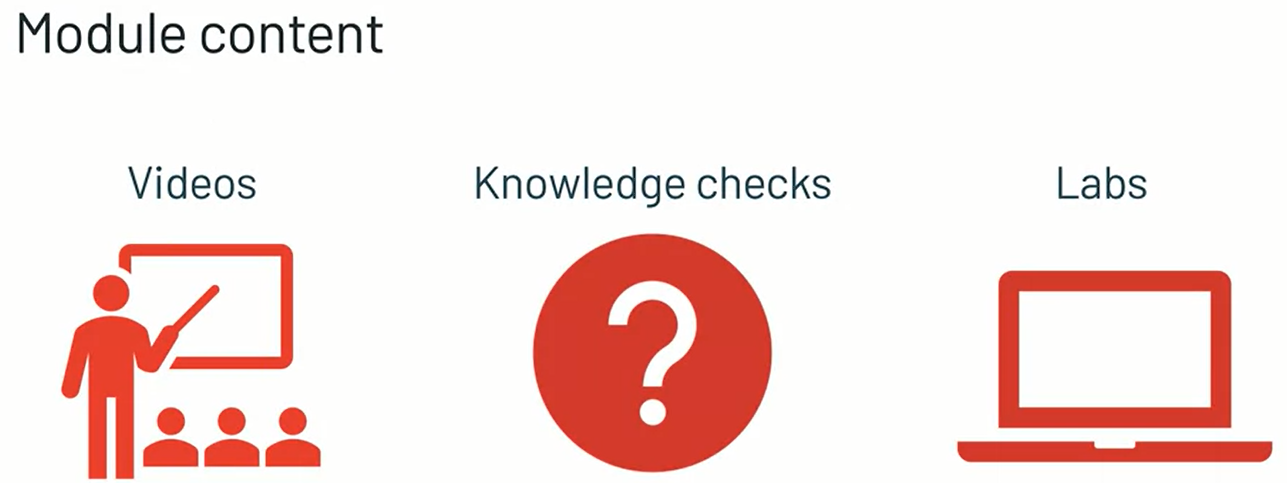
# **An Introduction to Statistics and Probability**

# **Module and Lesson Intro**

Hi, everyone, and welcome back to another module of Data Science Fundamentals for data analysts. [Herkese merhaba, veri analistleri için başka bir Veri Bilimi Temelleri modülüne tekrar hoş geldiniz.] In the previous module, you learned how data science is all about using interdisciplinary skills to solve real world problems. [Önceki modülde, veri biliminin gerçek dünya sorunlarını çözmek için disiplinler arası becerileri nasıl kullanmakla ilgili olduğunu öğrendiniz.] Next, we'll learn about probability and statistics. [Ardından, olasılık ve istatistik hakkında bilgi edineceğiz.]



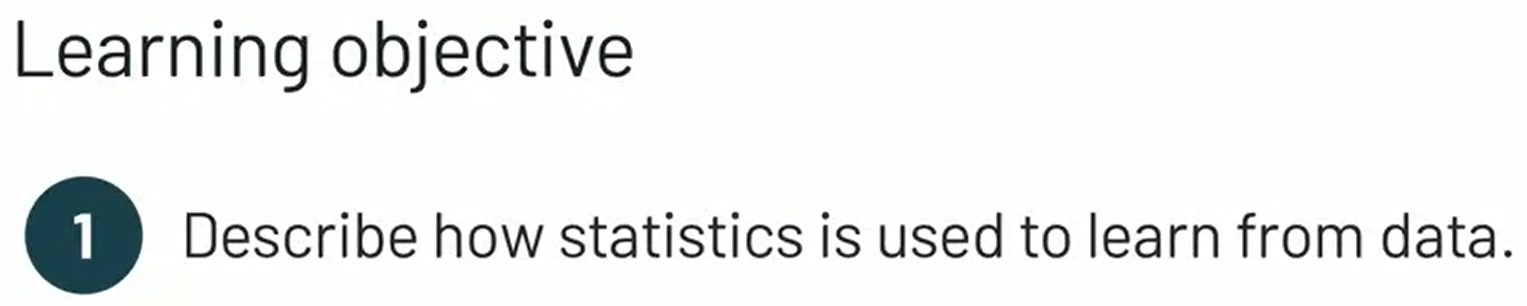
In this module, we have three lessons, and at the end of those three lessons, you'll be able to apply foundational data science concepts from inferential statistics. [Bu modülde üç dersimiz var ve bu üç dersin sonunda çıkarımsal istatistiklerden temel veri bilimi kavramlarını uygulayabileceksiniz.]



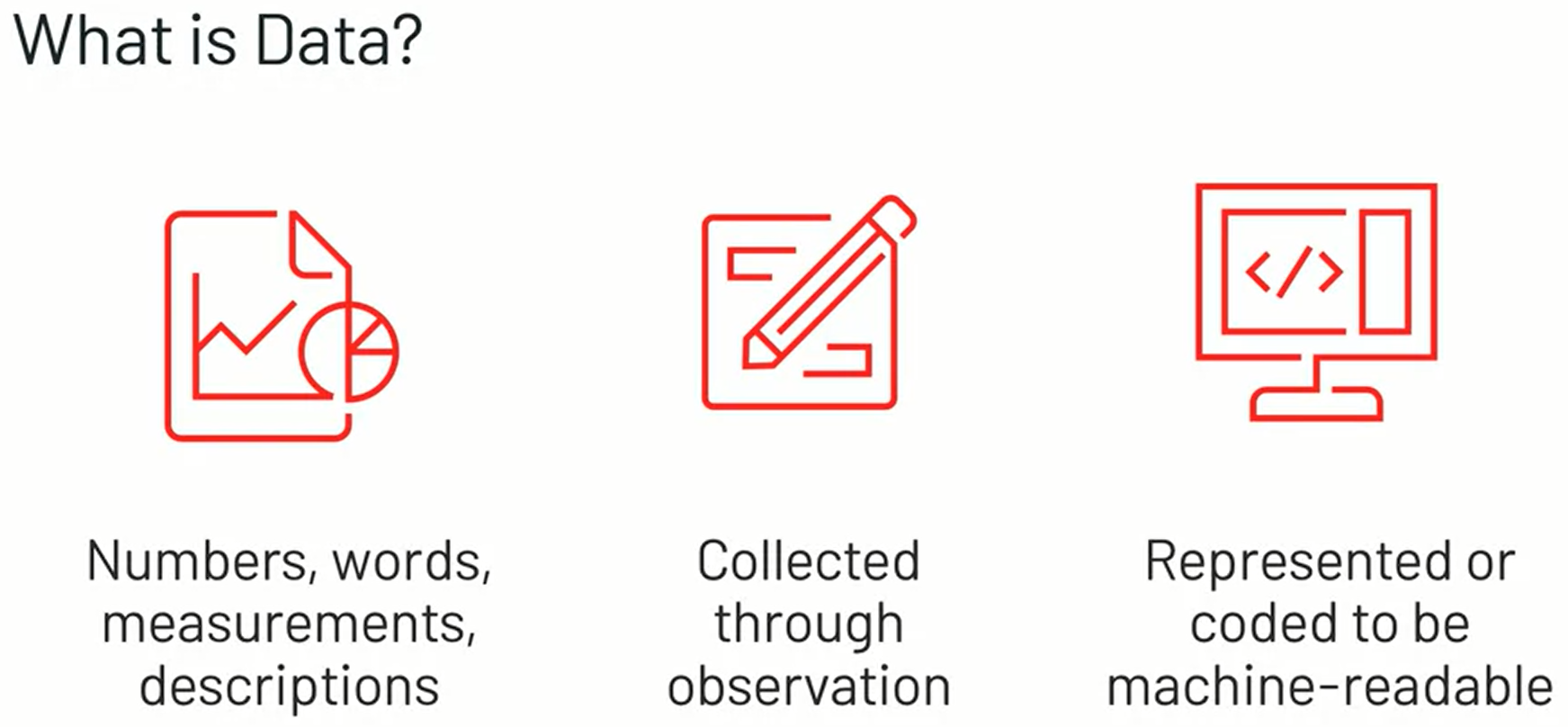
All of these concepts will be taught in a series of videos, knowledge checks, and labs. [Tüm bu kavramlar bir dizi video, bilgi kontrolleri ve laboratuvarlarda öğretilecektir.] Just like before, the videos will focus on concepts and the activities will allow you to apply what you've learned. [Daha önce olduğu gibi, videolar kavramlara odaklanacak ve etkinlikler öğrendiklerinizi uygulamanıza izin verecek.] Now, click "Next" to hear from another one of my wonderful colleagues at Databricks, Emma. [Şimdi, Databricks'teki harika meslektaşlarımdan biri olan Emma'dan haber almak için 'İleri'ye tıklayın.]

# **An Introduction to Statistics**

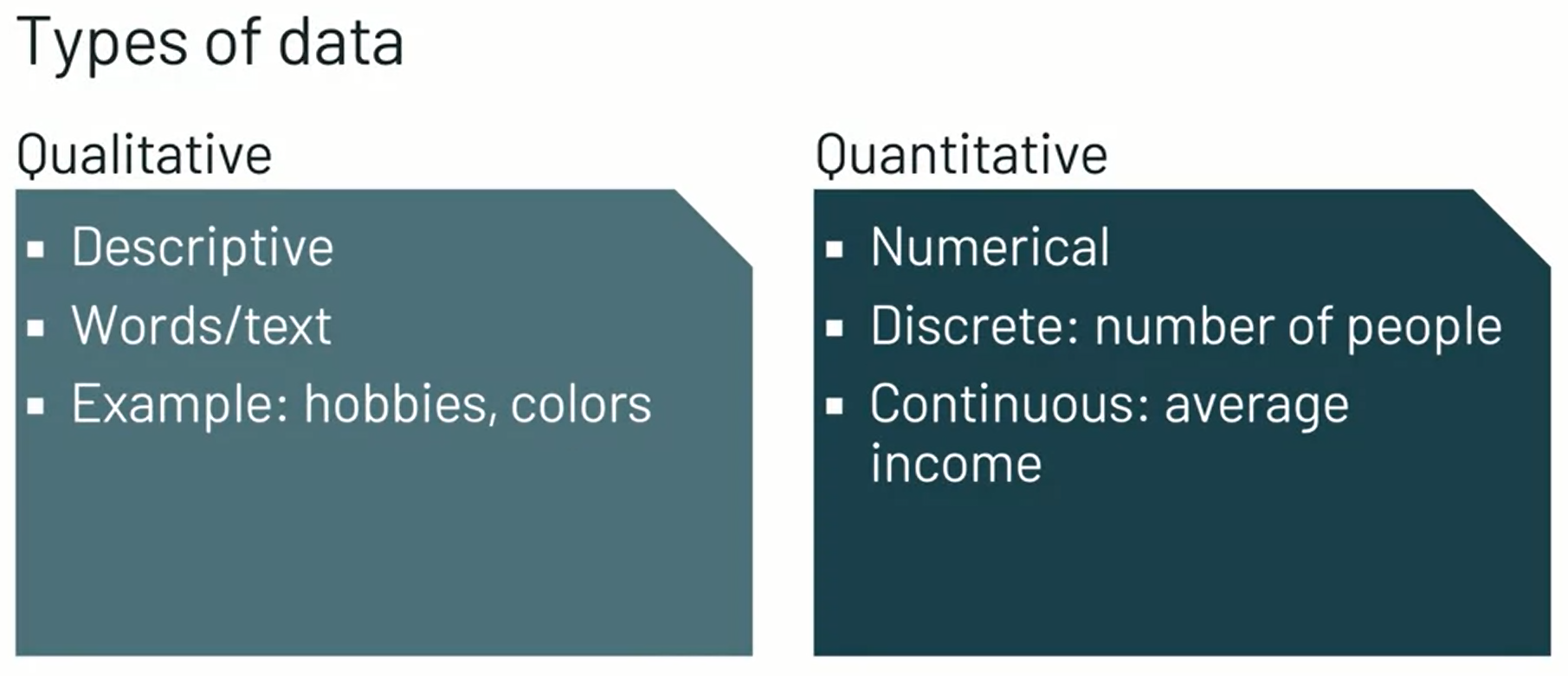
Hello, and welcome back. [Merhaba ve tekrar hoş geldiniz.] My name is Emma Freeman and I'm a Curriculum Engineer here at Databricks. [Benim adım Emma Freeman ve Databricks'te Müfredat Mühendisiyim.] I live in the San Francisco Bay Area and I enjoy surfing and land dancing. [San Francisco Körfez Bölgesi'nde yaşıyorum ve sörf yapmaktan ve kara dansı yapmaktan hoşlanıyorum.] Before Databricks, I was a data scientists at several companies, and I've also taught data Science and Python programming. [Databricks'ten önce birkaç şirkette veri bilimciydim ve ayrıca veri bilimi ve Python programlama dersleri verdim.] As we mentioned in the module introduction, statistics forms the foundation of data science, and it's a very important tool for learning from data. [Modül girişinde bahsettiğimiz gibi istatistik, veri biliminin temelini oluşturur ve verilerden öğrenmek için çok önemli bir araçtır.] The field of statistics is all about extracting meaning from data by finding patterns and relationships using mathematics. [İstatistik alanı, matematik kullanarak örüntüler ve ilişkiler bularak verilerden anlam çıkarmakla ilgilidir.] Once we extract this meaning, we can use it to help us make decisions. [Bu anlamı çıkardığımızda, karar vermemize yardımcı olması için kullanabiliriz.]



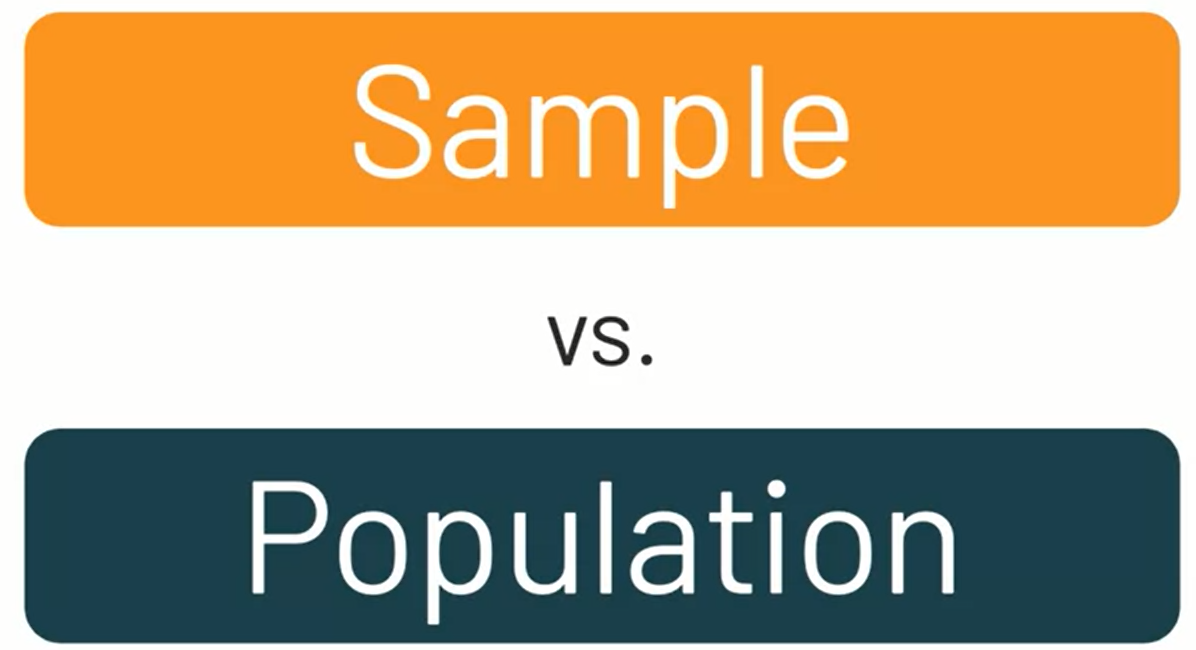
By the end of this video, you'll be able to describe how statistics is used to learn from data. [Bu videonun sonunda, verilerden öğrenmek için istatistiklerin nasıl kullanıldığını açıklayabileceksiniz.]



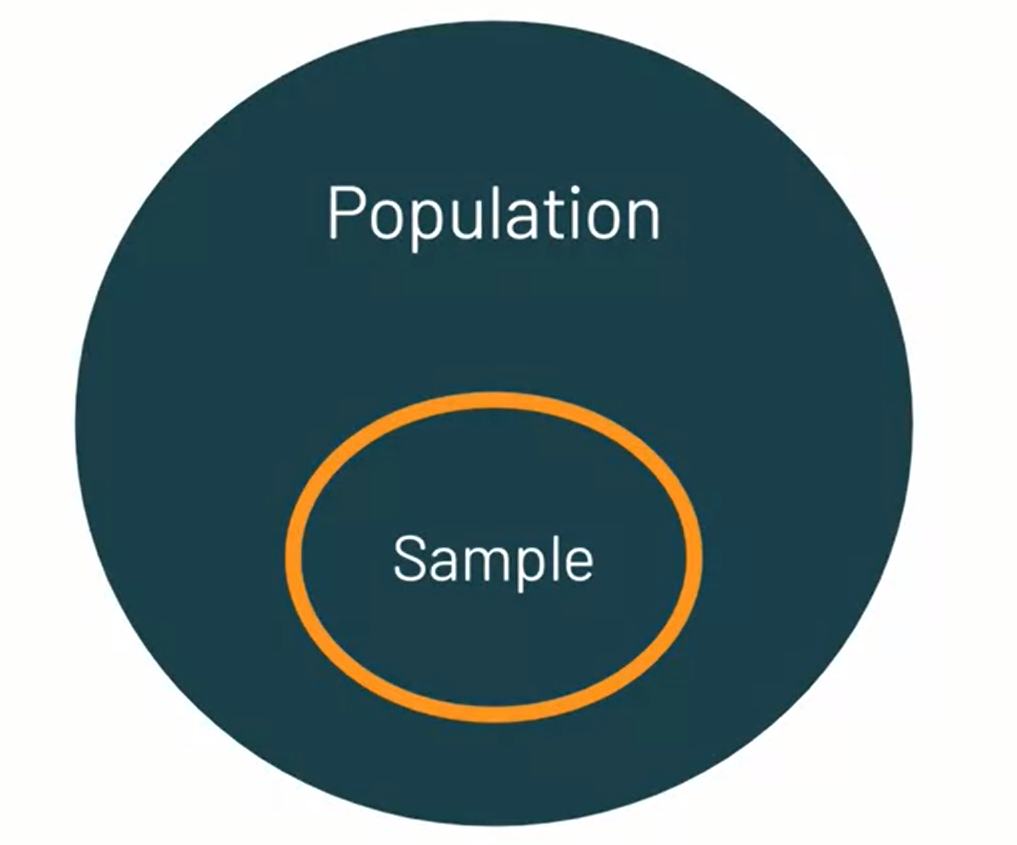
Before we dive into statistics, let's talk about data. [İstatistiklere geçmeden önce verilerden bahsedelim.] What exactly is data? [Veri tam olarak nedir?] Data can be a collection of information like numbers, words, measurements, or descriptions that's been collected through observation of something. [Veriler, bir şeyin gözlemlenmesi yoluyla toplanan sayılar, kelimeler, ölçümler veya açıklamalar gibi bir bilgi koleksiyonu olabilir.] There are several ways that data can be collected. [Veri toplamanın birkaç yolu vardır.] We can collect data through observation. [Gözlem yoluyla veri toplayabiliriz.] For example, counting the number of fish that swim past a certain point in a lake over a 10 minute period. [Örneğin, 10 dakikalık bir süre içinde bir gölde belirli bir noktadan yüzerek geçen balık sayısını saymak.] Another common way of collecting data is through a survey. [Veri toplamanın diğer bir yaygın yolu da ankettir.] In order for information to be considered data, it has to be represented or coded in some form that is suitable for processing. [Bilginin veri olarak kabul edilebilmesi için, işlenmeye uygun bir biçimde temsil edilmesi veya kodlanması gerekir.]



There are many different types of data. [Birçok farklı veri türü vardır.] At the highest level, data can be defined as either qualitative or quantitative. [En üst düzeyde, veriler nitel veya nicel olarak tanımlanabilir.] Qualitative data is descriptive. [Nitel veriler tanımlayıcıdır.] It describes the quality of the thing being observed. [Gözlenen şeyin kalitesini tanımlar.] Some examples of qualitative data are the most popular hobbies of people living in the United States, or the colors of fish that live in Bass lake. [Nitel verilere bazı örnekler, Amerika Birleşik Devletleri'nde yaşayan insanların en popüler hobileri veya Bass gölünde yaşayan balıkların renkleridir.] Quantitative data is numerical. [Nicel veriler sayısaldır.] It gives us a specific quantity. [Bize belirli bir miktar verir.] Furthermore, quantity of data can be discrete or continuous. [Ayrıca, veri miktarı ayrık veya sürekli olabilir.] Discrete data can only take certain specified values such as whole numbers. [Ayrık veriler, yalnızca tam sayılar gibi belirli belirli değerleri alabilir.] It represents a count of something. [Bir şeyin sayısını temsil eder.] For example, a count of the number of people within different age ranges who live in the United States, or how many species of fish live in Bass lake. [Örneğin, Amerika Birleşik Devletleri'nde yaşayan farklı yaş aralıklarındaki insan sayısı veya Bass gölünde kaç balık türü yaşıyor.] Continuous data can take any value within some range. [Sürekli veriler belirli bir aralıkta herhangi bir değer alabilir.] It represents a measurement. [Bir ölçümü temsil eder.] For example, the average income of adults in the US, or the length in centimeters of fishes found in Bass lake. [Örneğin, ABD'deki yetişkinlerin ortalama geliri veya Bass gölünde bulunan balıkların santimetre cinsinden uzunluğu.]



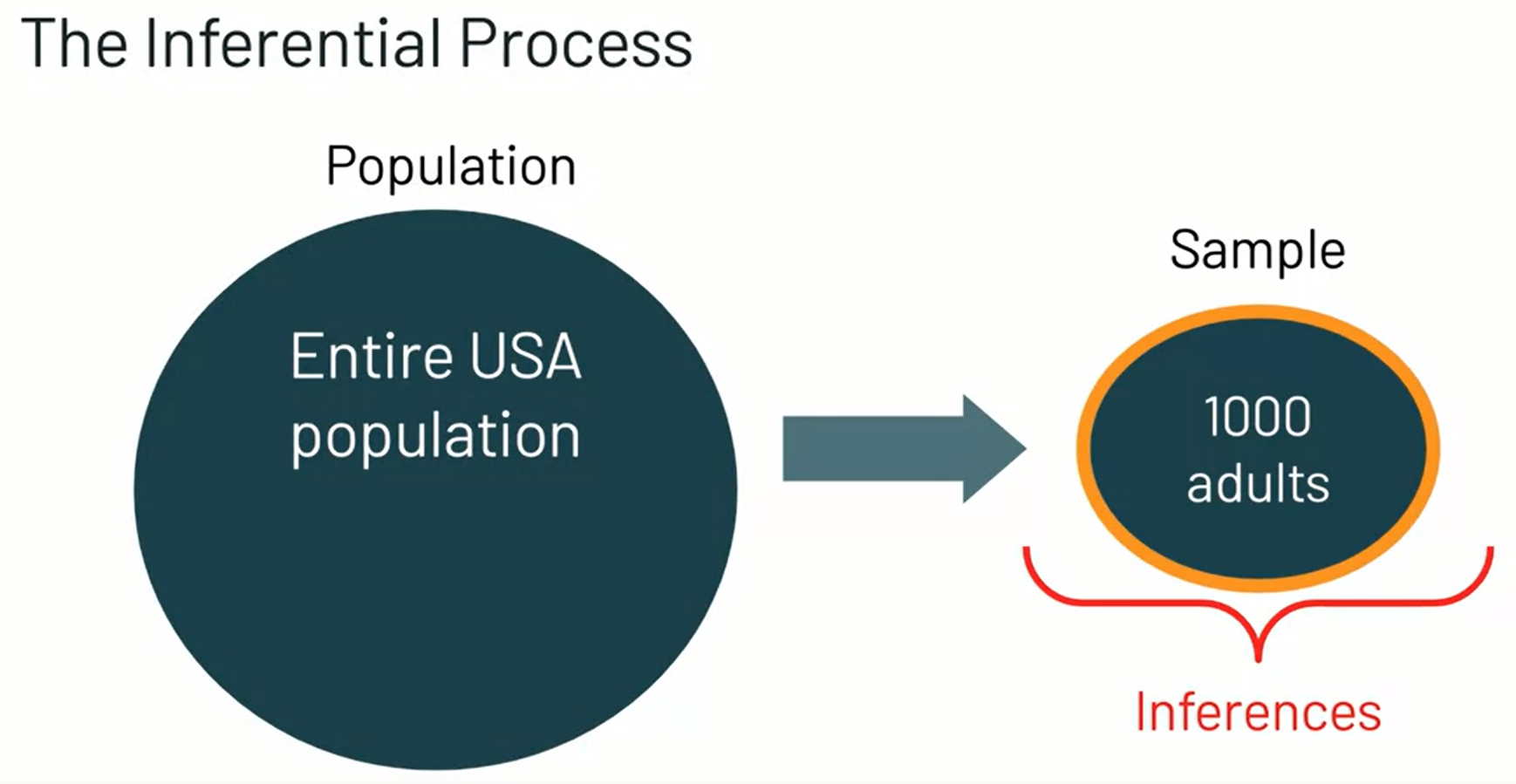
In addition to these main categories of types of data, we also have the concept of a sample versus an entire population. [Bu ana veri türleri kategorilerine ek olarak, tüm popülasyona karşı bir örnek kavramına da sahibiz.] This concept is central to statistics and the foundation of drawing conclusions from data. [Bu kavram, istatistiklerin ve verilerden sonuç çıkarmanın temelidir.]



A population is the entire group of something. [Bir popülasyon, bir şeyin tüm grubudur.] A sample is a part or subset of an entire population. [Örnek, tüm popülasyonun bir parçası veya alt kümesidir.]



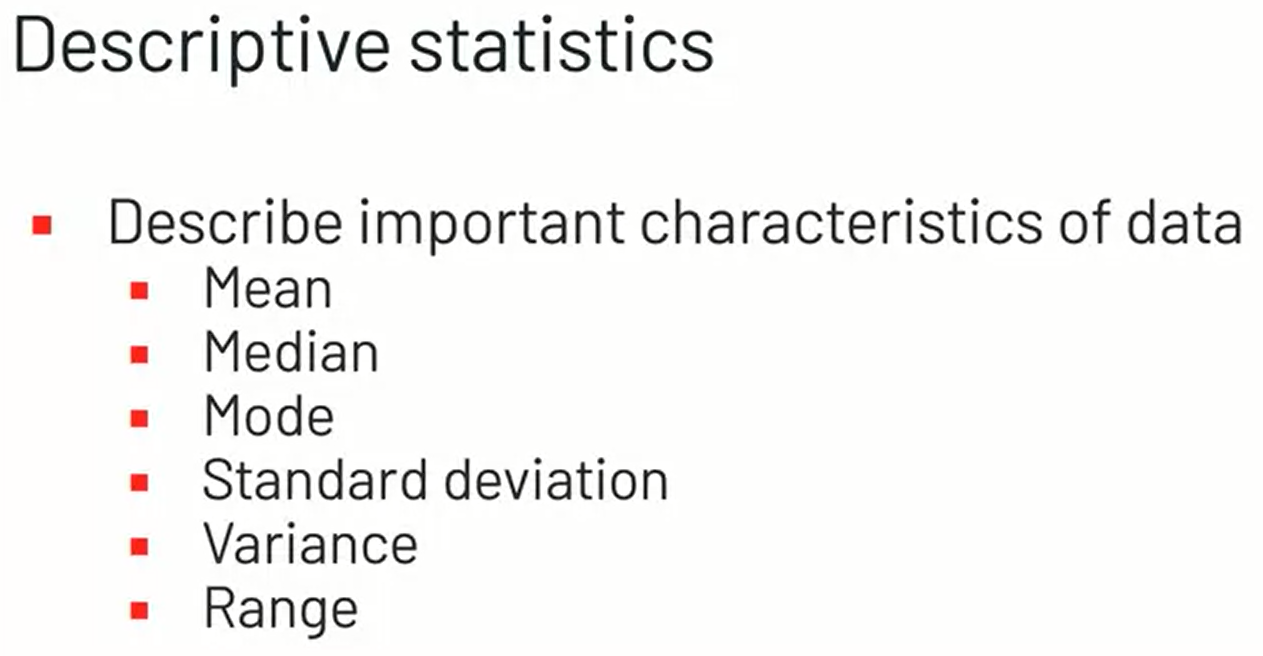
For example, a population could refer to the entire adult population of the United States, or the entire population of fish in Bass lake. [Örneğin, bir popülasyon, Amerika Birleşik Devletleri'ndeki tüm yetişkin popülasyonu veya Bass gölündeki tüm balık popülasyonunu ifade edebilir.] But in statistical terms, population doesn't just refer to a population of people or animals or things like that. [Ancak istatistiksel olarak, nüfus sadece bir insan veya hayvan popülasyonunu veya bunun gibi şeyleri ifade etmez.] Statisticians also talk about populations of events, procedures, things or observations. [İstatistikçiler ayrıca olay popülasyonları, prosedürler, şeyler veya gözlemler hakkında konuşurlar.] From our previous example, if we were going to survey 1,000 adults in the United States about their hobbies, or maybe observe only the fish in a certain part of Bass lake for a specified period of time. [Bir önceki örneğimizden, Amerika Birleşik Devletleri'ndeki 1.000 yetişkine hobileri hakkında anket yapacaksak ya da belki sadece Bass gölünün belirli bir bölümünde belirli bir süre boyunca balıkları gözlemleyeceksek.] This would be a sample of the population. [Bu, popülasyonun bir örneği olacaktır.]



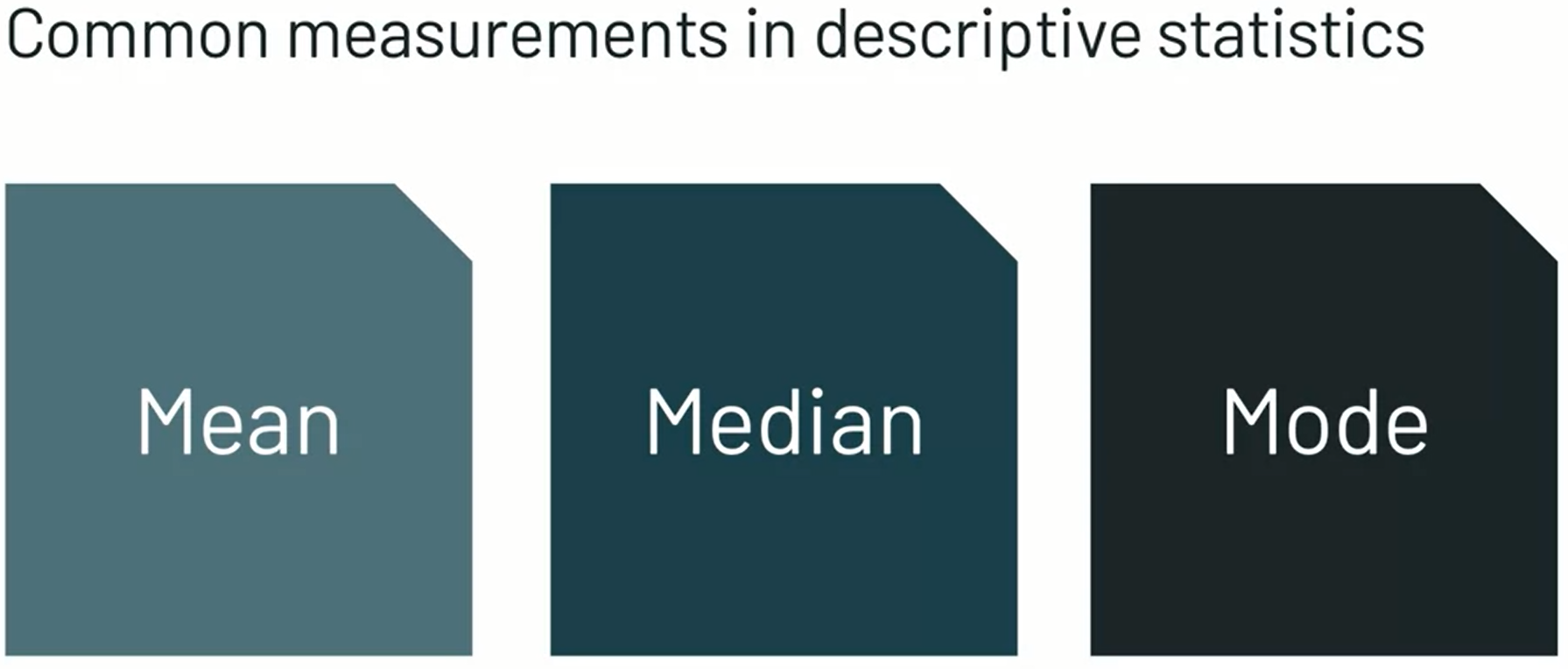
Usually, it would be impossible, too costly, or too difficult to collect data for an entire population. [Genellikle, tüm popülasyon için veri toplamak imkansız, çok maliyetli veya çok zor olurdu.] Imagine trying to survey all 209 million adults in the US about their hobbies. [ABD'deki 209 milyon yetişkinin tümüne hobileri hakkında anket yapmaya çalıştığınızı hayal edin.] This would take an extremely long time. [Bu son derece uzun bir zaman alacaktır.] Not to mention, it would be impossible to locate contact and get responses from everyone. [Bahsetmemek gerekirse, muhatap bulmak ve herkesten yanıt almak imkansız olurdu.] Instead, if we wanted to find out what are the most popular hobbies of adults in the US, we would take a sample of the whole population and do a survey about their hobbies. [Bunun yerine, ABD'deki yetişkinlerin en popüler hobilerinin neler olduğunu öğrenmek isteseydik, tüm popülasyondan bir örnek alır ve hobileri hakkında bir anket yapardık.] From this data, we can make inferences. [Bu verilerden çıkarımlarda bulunabiliriz.] That is, draw conclusions about the population as a whole. [Yani, bir bütün olarak nüfus hakkında sonuçlar çıkarın.] These are the general concepts that form the foundation of statistical data analysis. [Bunlar istatistiksel veri analizinin temelini oluşturan genel kavramlardır.] In the next video, we'll dive deeper into one of the main types of statistics. [Bir sonraki videoda, ana istatistik türlerinden birine daha derinden dalacağız.]

# **Descriptive Statistics**

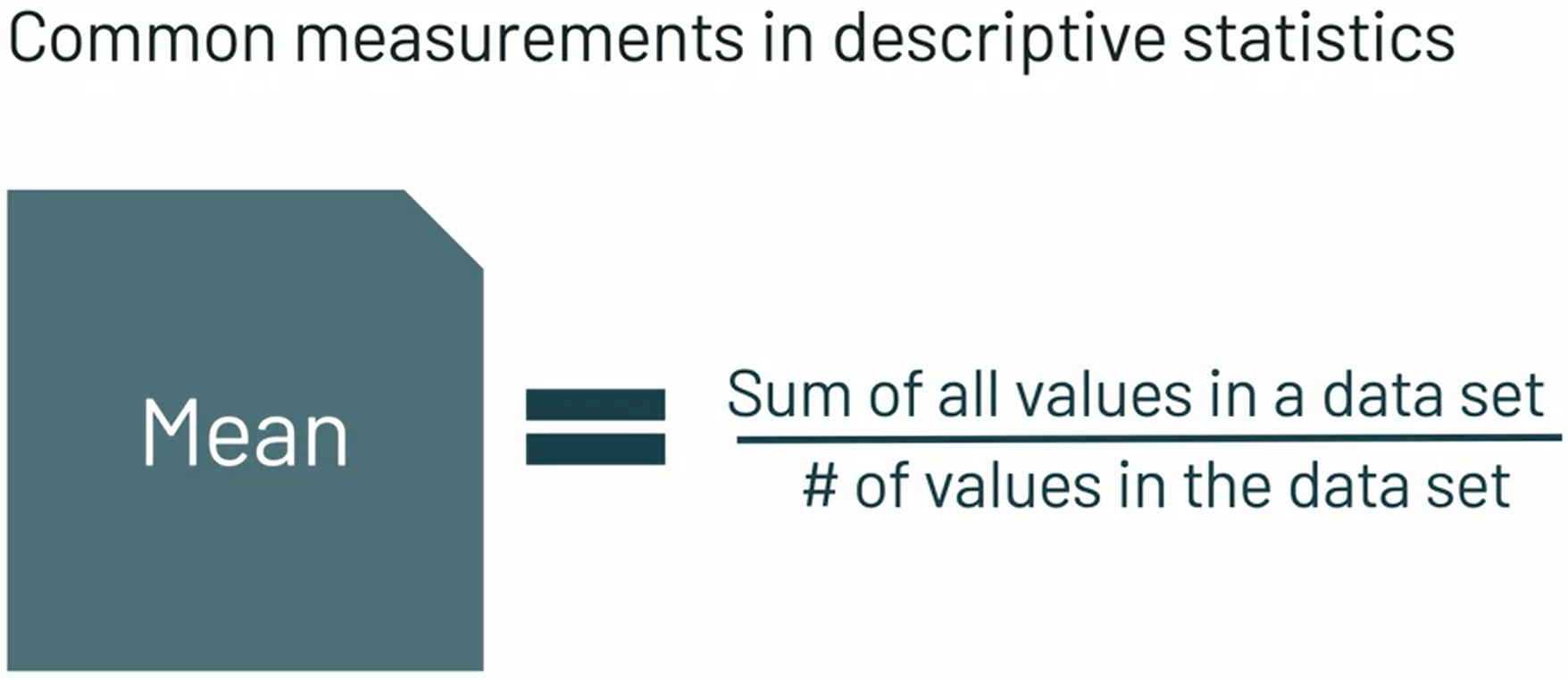
In the previous video, we introduced you to some key terms and concepts of statistics. [Bir önceki videoda, size istatistikle ilgili bazı temel terimleri ve kavramları tanıttık.] In this video, we'll learn about one of the two main categories of statistics; descriptive statistics. [Bu videoda, iki ana istatistik kategorisinden birini öğreneceğiz; tanımlayıcı istatistikler.] Like I mentioned, there's two main branches of statistics; descriptive and inferential. [Bahsettiğim gibi, istatistiğin iki ana dalı vardır; açıklayıcı ve çıkarımsal.] Each of these provides a different type of insights from data, and both of them used together can give us a more complete picture. [Bunların her biri verilerden farklı türde içgörüler sağlar ve ikisi birlikte kullanıldığında bize daha eksiksiz bir resim verebilir.]



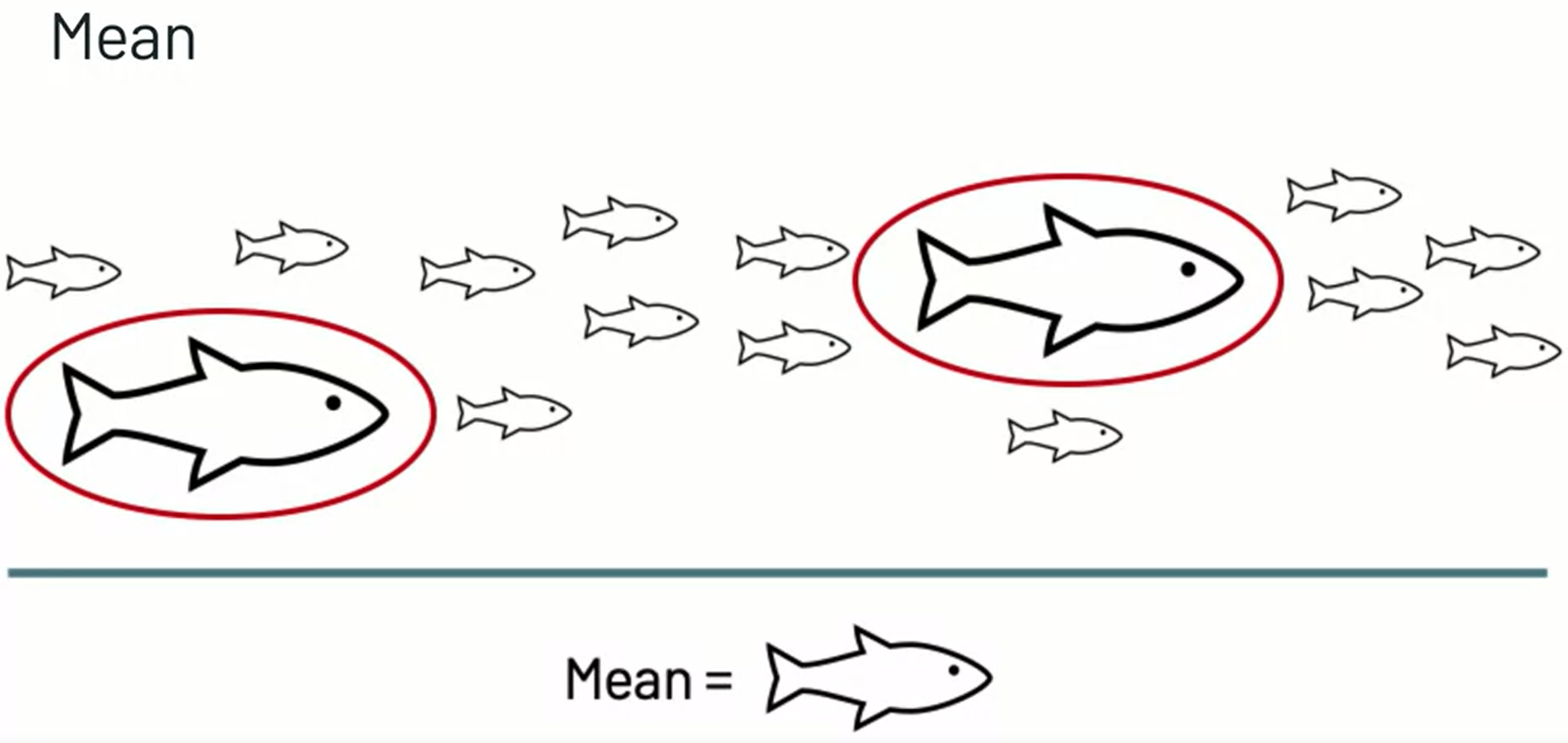
Descriptive statistics describe important characteristics of data, which we can summarize and display in charts, tables and graphs. [Tanımlayıcı istatistikler, çizelgeler, tablolar ve grafiklerde özetleyebileceğimiz ve görüntüleyebileceğimiz verilerin önemli özelliklerini tanımlar.] This includes measures of central tendency such as mean, median and mode, and measures of dispersion such as standard deviation, variance and range. [Bu, ortalama, medyan ve mod gibi merkezi eğilim ölçülerini ve standart sapma, varyans ve aralık gibi dağılım ölçülerini içerir.] Don't worry if you're not familiar with these terms yet, we'll define and describe them in the next slides. [Henüz bu terimlere aşina değilseniz endişelenmeyin, bunları sonraki slaytlarda tanımlayıp açıklayacağız.]



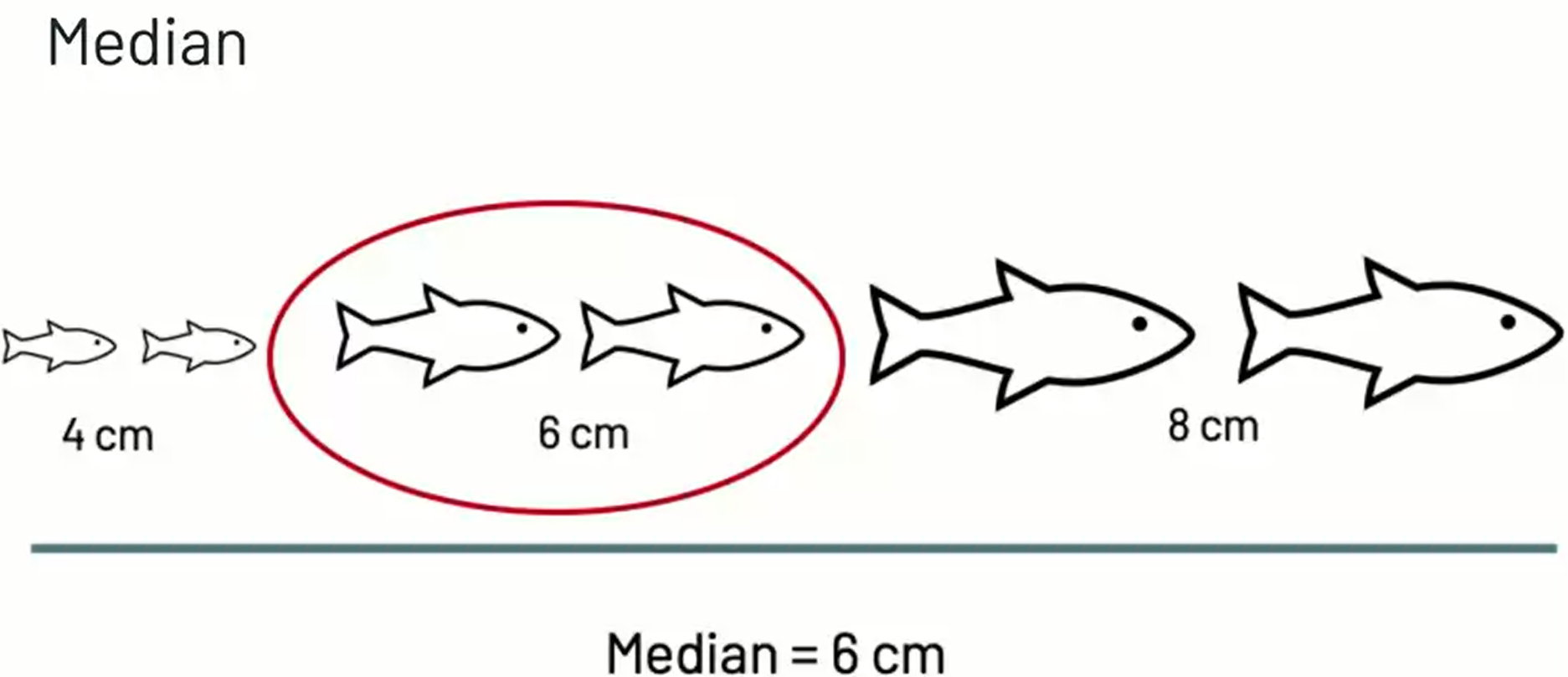
Measures of central tendency are also referred to as Summary Statistics. [Merkezi eğilim ölçüleri ayrıca Özet İstatistik olarak da adlandırılır.] Each measure is a single value that attempts to describe a sample by identifying the central position within that sample. [Her ölçü, o örnek içindeki merkezi konumu belirleyerek bir örneği tanımlamaya çalışan tek bir değerdir.] The mean, median, and mode are all measures of central tendency. [Ortalama, medyan ve mod, tümü merkezi eğilim ölçüleridir.]



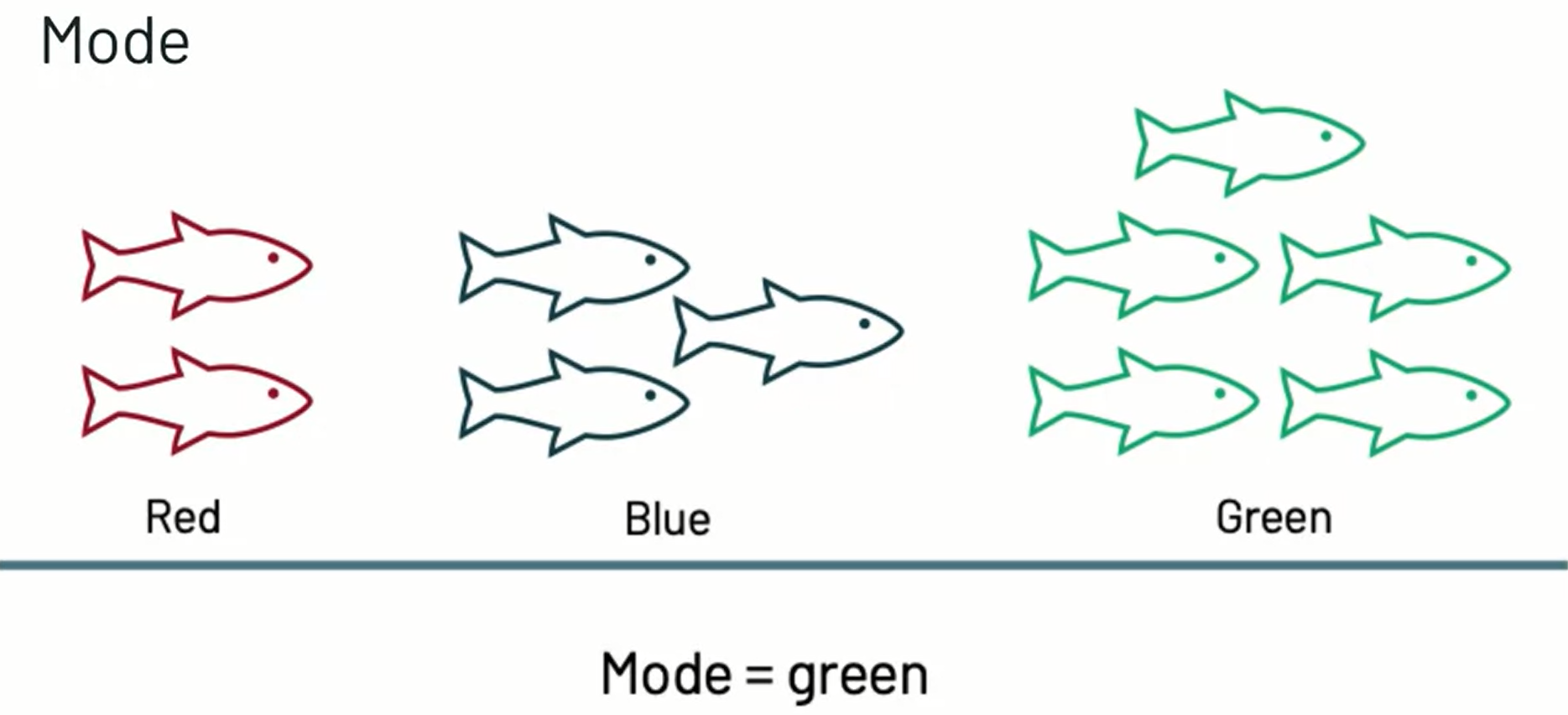
The mean, commonly known as the average, is the most popular and well-known measure of central tendency. [Genel olarak ortalama olarak bilinen ortalama, en popüler ve en iyi bilinen merkezi eğilim ölçüsüdür.] It's equal to the sum of all the values in the dataset, divided by the number of values in the dataset. [Veri kümesindeki tüm değerlerin toplamının veri kümesindeki değer sayısına bölünmesine eşittir.] So if we have n values in a dataset and they have the values X1, X2, and so on, the sample mean, usually denoted by x bar, is written as shown in the formula on the screen. [Yani bir veri setinde n değerimiz varsa ve X1, X2 vb. değerlere sahiplerse, genellikle x çubuğu ile gösterilen örnek ortalama, ekrandaki formülde gösterildiği gibi yazılır.]



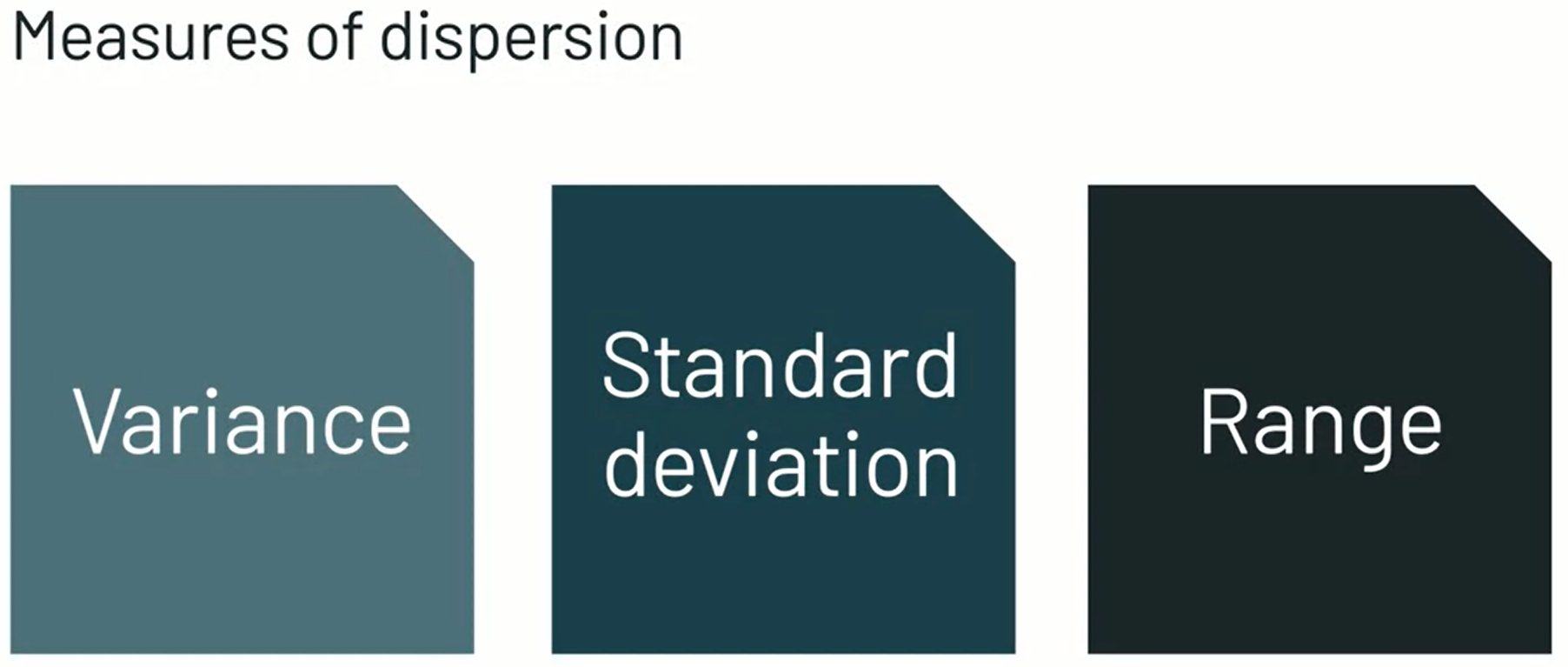
One thing to consider when using the mean as a measure of central tendency is that it's very sensitive to outliers, which are unusually small or large values. [Ortalamayı bir merkezi eğilim ölçüsü olarak kullanırken göz önünde bulundurulması gereken bir şey, alışılmadık derecede küçük veya büyük değerler olan aykırı değerlere karşı çok hassas olmasıdır.] For example, if most of the fish from our bass lake sample are less than eight centimeters long, but there are a few fish that are twice as big, our mean is going to be a skewed by those large values and it won't accurately represent the actual sample. [Örneğin, levrek gölü örneğimizdeki balıkların çoğu sekiz santimetreden daha kısaysa, ancak iki katı büyüklükte birkaç balık varsa, ortalamamız bu büyük değerlerle çarpık olacaktır ve bu doğru olmayacaktır. gerçek örneği temsil eder.]



One way to deal with this is to instead, use the median. [Bununla başa çıkmanın bir yolu, bunun yerine medyanı kullanmaktır.] The median is the middle value in our dataset if we arrange all of the values in order. [Tüm değerleri sırayla düzenlersek, medyan veri setimizin ortasındaki değerdir.] As shown on the screen, the median value for fish size is 6 centimeters. [Ekranda görüldüğü gibi balık boyutu için ortanca değer 6 santimetredir.]



The mode is simply the most frequent value in the dataset, and it's often used to describe categorical data. [Mod, veri kümesindeki en sık görülen değerdir ve genellikle kategorik verileri tanımlamak için kullanılır.] In this case, the mode represents the most common or popular category. [Bu durumda, mod en yaygın veya popüler kategoriyi temsil eder.] For example, if there are three colors of fish and bass lake, and we observe that most of the fish are green, then green would be the mode for the color variable. [Örneğin, balık ve levrek gölünün üç rengi varsa ve balıkların çoğunun yeşil olduğunu gözlemlersek, renk değişkeni için mod yeşil olacaktır.]



In addition to measures of central tendency, we also have measures of dispersion. [Merkezi eğilim ölçülerine ek olarak, dağılım ölçülerimiz de var.] These measures represent how spread out a dataset is. [Bu önlemler, bir veri kümesinin ne kadar yayılmış olduğunu gösterir.]



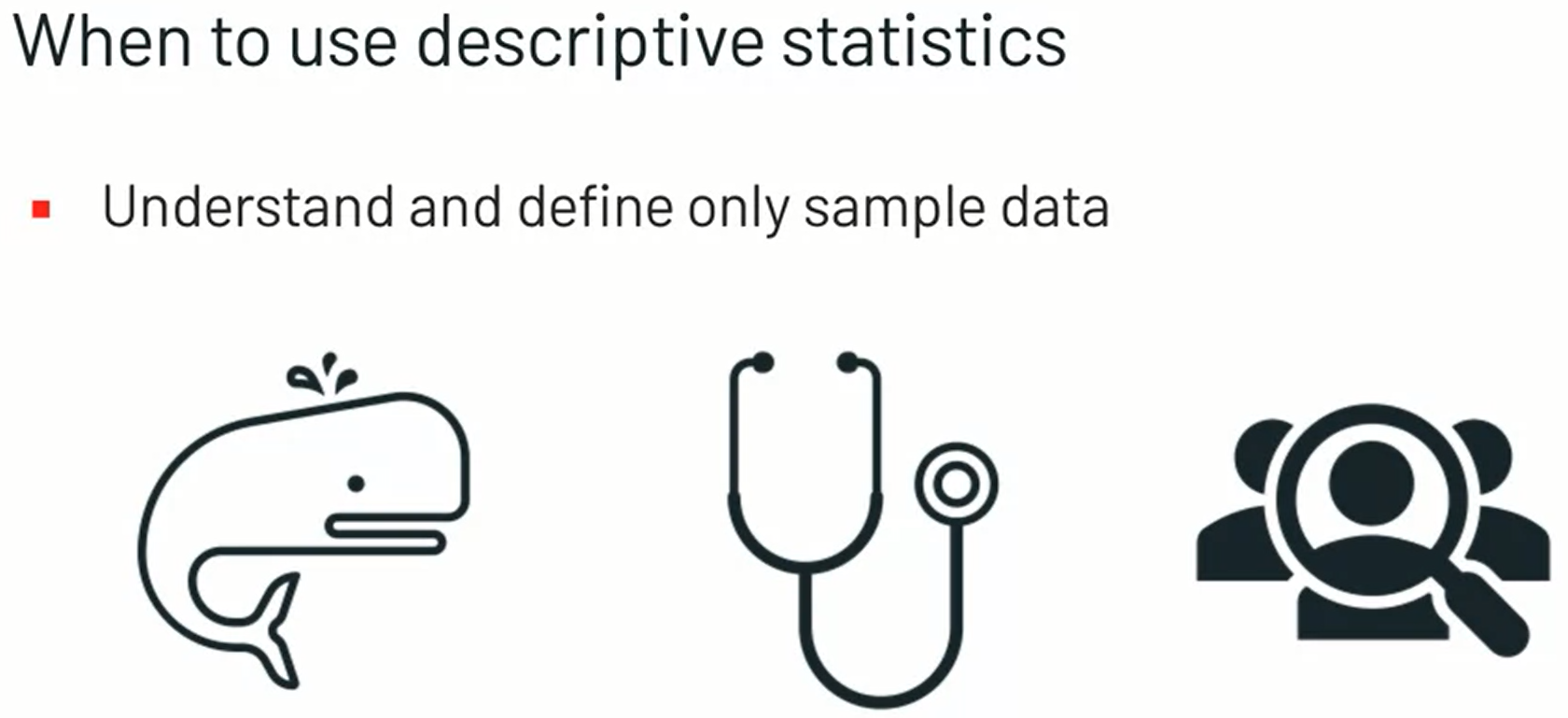
The variance is the average of the squared difference of each data point from the mean. [Varyans, her bir veri noktasının ortalamadan karesi alınmış farkının ortalamasıdır.] That might not sound like it makes a lot of sense, and this number in itself isn't actually very useful, which is why we mostly use it to compute the standard deviation. [Bu kulağa pek mantıklı gelmeyebilir ve bu sayının kendisi aslında çok kullanışlı değil, bu yüzden çoğunlukla standart sapmayı hesaplamak için kullanıyoruz.]



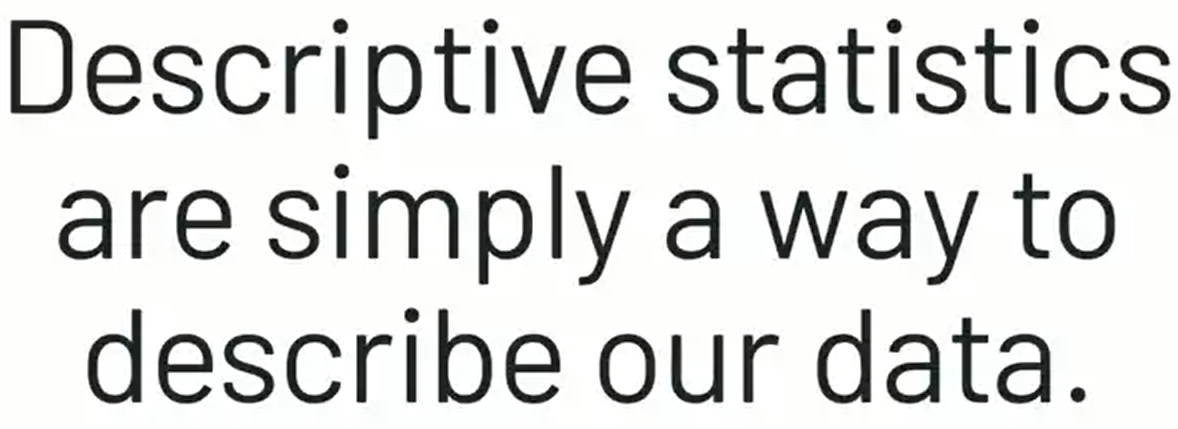
Standard deviation is the most common measure of dispersion, and it tells us how much our data is spread out around the mean. [Standart sapma, en yaygın dağılım ölçüsüdür ve bize verilerimizin ortalamaya ne kadar yayıldığını söyler.] When the standard deviation is larger, it means that our data is more spread out. [Standart sapma daha büyük olduğunda, verilerimizin daha fazla yayıldığı anlamına gelir.] When it's smaller, it means that more of our data points are closer to the mean. [Daha küçük olduğunda, veri noktalarımızın çoğunun ortalamaya daha yakın olduğu anlamına gelir.]



The range of a dataset is the difference between the largest and the smallest observation in the data. [Bir veri kümesinin aralığı, verilerdeki en büyük ve en küçük gözlem arasındaki farktır.] While this is easy to calculate, it's very sensitive to outliers and it doesn't really tell us a lot about our dataset. [Bunun hesaplanması kolay olsa da, aykırı değerlere karşı çok hassastır ve bize veri kümemiz hakkında pek bir şey söylemez.] It only says what the largest and smallest values are. [Yalnızca en büyük ve en küçük değerlerin ne olduğunu söyler.] The inter-quartile range, on the other hand, describes the difference between the 25th and 75th percentile of the data. [Çeyrekler arası aralık ise, verilerin 25. ve 75. yüzdelikleri arasındaki farkı tanımlar.] In other words, the middle 50 percent, and that can be more useful in understanding the distribution of data. [Başka bir deyişle, orta yüzde 50 ve bu, verilerin dağılımını anlamada daha yararlı olabilir.]



We use descriptive statistics when we want to understand and define only our sample data. [Tanımlayıcı istatistikleri yalnızca örnek verilerimizi anlamak ve tanımlamak istediğimizde kullanırız.] We can summarize and describe the sample using known quantitative measurements. [Bilinen nicel ölçümleri kullanarak örneği özetleyebilir ve tanımlayabiliriz.] This process of exploring a dataset using descriptive statistics is called exploratory data analysis, otherwise known as EDA, and is a big part of the data science workflow. [Tanımlayıcı istatistikleri kullanarak bir veri kümesini keşfetme sürecine, aksi takdirde EDA olarak bilinen keşifsel veri analizi denir ve veri bilimi iş akışının büyük bir parçasıdır.] Some examples of where we might see descriptive statistics used are in scientific or medical research to describe the study or the sample population. [Tanımlayıcı istatistiklerin kullanıldığını görebileceğimiz bazı örnekler, çalışmayı veya örnek popülasyonu tanımlamak için bilimsel veya tıbbi araştırmalardadır.] For example, we might want to look at the mean and median number of days that gray whales take to complete their annual migration. [Örneğin, gri balinaların yıllık göçlerini tamamlamak için geçirdikleri ortalama gün ve medyan gün sayısına bakmak isteyebiliriz.] Or maybe we want to look at the standard deviation of patients' blood pressure in a clinical trial. [Ya da bir klinik deneyde hastaların kan basıncının standart sapmasına bakmak isteyebiliriz.] We also often see descriptive statistics used in business applications, like observing certain characteristics of customer behavior. [Ayrıca, müşteri davranışının belirli özelliklerini gözlemlemek gibi iş uygulamalarında kullanılan tanımlayıcı istatistikleri de sıklıkla görüyoruz.]



Descriptive statistics don't allow us to make conclusions beyond the data we've analyzed or reach conclusions regarding any hypothesis we might have made. [Tanımlayıcı istatistikler, analiz ettiğimiz verilerin ötesinde sonuçlar çıkarmamıza veya yapmış olabileceğimiz herhangi bir hipotezle ilgili sonuçlara varmamıza izin vermez.] They are simply a way to describe our sample dataset. [Bunlar sadece örnek veri setimizi tanımlamanın bir yoludur.]

# **Descriptive Statistics Lab Intro**

It's time for another lab. In the upcoming lab, you will get some hands-on practice with gathering descriptive statistics about a dataset. In this lab, you will complete a series of guided descriptive statistics exercises. Good luck.

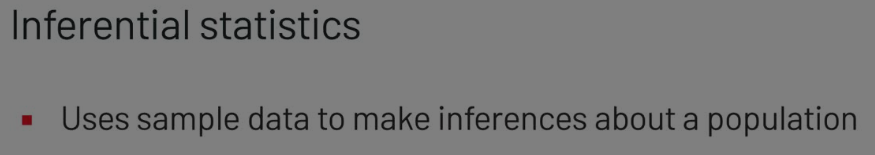
# **Descriptive Statistics Lab**

Next, you'll complete a lab on descriptive statistics using Databricks.

To do so, log in to your Databricks Community Edition account and locate your DSFDA folder. Then, open Module 3, then click on "3.1.1 Lab - Descriptive Statistics" to open your notebook.

If you haven’t already imported the course DBC file, review instructions in Module 1's "Hands-on with Databricks".

# **Inferential Statistics**



In the previous video you learned about how you can use descriptive statistics to describe a data sample. [Önceki videoda, bir veri örneğini tanımlamak için tanımlayıcı istatistikleri nasıl kullanabileceğinizi öğrendiniz.] In this video will learn about the other main category of statistics, inferential statistics. [Bu videoda istatistiklerin diğer ana kategorisi olan çıkarımsal istatistikler hakkında bilgi edineceksiniz.] Inferential statistics uses sample data to make inferences about the population. [Çıkarımsal istatistikler, popülasyon hakkında çıkarımlar yapmak için örnek verileri kullanır.] It's a way of drawing conclusions from our sample data and then generalizing these conclusions to a larger population. [Örnek verilerimizden sonuçlar çıkarmanın ve daha sonra bu sonuçları daha büyük bir popülasyona genellemenin bir yolu.] Since we can't usually study an entire population, we collect a sample that's representative of that entire population, and then calculate measures about that sample. [Genellikle tüm popülasyonu inceleyemeyeceğimiz için, tüm popülasyonu temsil eden bir örneklem toplarız ve sonra o örnekle ilgili ölçümleri hesaplarız.] We then use the sample characteristics to make generalizations or inferences about the population. [Daha sonra popülasyon hakkında genellemeler veya çıkarımlar yapmak için örnek özelliklerini kullanırız.] In order to make accurate inferences, the sample has to be representative of the actual population. [Doğru çıkarımlar yapabilmek için örneklemin gerçek popülasyonu temsil etmesi gerekir.] A representative sample is one where each and every member of that population has an equal and mutually exclusive chance of being selected as part of our sample. [Temsili bir örneklem, o popülasyonun her bir üyesinin, örneklemimizin bir parçası olarak seçilmek için eşit ve birbirini dışlayan bir şansa sahip olduğu bir örnektir.] This means that we're not biased and how the sample is chosen from the population. [Bu, önyargılı olmadığımız ve örneğin popülasyondan nasıl seçildiği anlamına gelir.] We wouldn't want to choose only the adults that we know in our town for our hobby survey, because this wouldn't reflect the true nature of the whole population and their hobbies. [Hobi anketimiz için sadece şehrimizde tanıdığımız yetişkinleri seçmek istemeyiz, çünkü bu, tüm nüfusun ve hobilerinin gerçek doğasını yansıtmaz.] Maybe everyone I know in Oakland, California is really into kayaking, but that doesn't mean that I can say that's true for everyone in the entire United States. [Belki Oakland, California'da tanıdığım herkes kanoyla ilgileniyor ama bu, bunun Birleşik Devletler'deki herkes için geçerli olduğunu söyleyebileceğim anlamına gelmiyor.] There are many sampling methods that are used to ensure that samples are representative and unbiased. [Örneklerin temsili ve tarafsız olmasını sağlamak için kullanılan birçok örnekleme yöntemi vardır.] I'm not going to go into the details of sampling methods here, it's just important to remember that selecting a good sample is critical for making inferences about a population. [Burada örnekleme yöntemlerinin ayrıntılarına girmeyeceğim, sadece iyi bir örnek seçmenin bir popülasyon hakkında çıkarımlarda bulunmak için kritik olduğunu hatırlamak önemlidir.] Despite our best efforts at making sure samples are representative of the population, samples would never be expected to perfectly represent the population. [Örneklerin popülasyonu temsil etmesini sağlamaya yönelik tüm çabalarımıza rağmen, örneklerin hiçbir zaman popülasyonu tam olarak temsil etmesi beklenemez.] This is where inferential statistics comes in. [Çıkarımsal istatistiklerin devreye girdiği yer burasıdır.] We can use them to better understand these errors. [Bu hataları daha iyi anlamak için bunları kullanabiliriz.] In turn, this allows us to determine whether the patterns we observe in our sample data actually generalized to the population, and whether these sample statistics adequately support our hypothesis about population. [Buna karşılık, bu, örnek verilerimizde gözlemlediğimiz modellerin gerçekten popülasyona genellenip genelleştirilmediğini ve bu örnek istatistiklerin popülasyon hakkındaki hipotezimizi yeterince destekleyip desteklemediğini belirlememize olanak tanır.] Now that you've learned about descriptive and inferential statistics, you might be wondering when you would use one or the other. [Artık tanımlayıcı ve çıkarımsal istatistikleri öğrendiğinize göre, birini veya diğerini ne zaman kullanacağınızı merak ediyor olabilirsiniz.] A good way to decide is to ask yourself whether you're trying to just describe a sample, or if you're trying to prove or disprove a hypothesis, or make some sort of prediction about the actual population. [Karar vermenin iyi bir yolu, kendinize sadece bir örneği tanımlamaya mı yoksa bir hipotezi kanıtlamaya mı yoksa çürütmeye mi yoksa gerçek popülasyon hakkında bir tür tahminde bulunmaya mı çalıştığınızı kendinize sormaktır.] To go back to our fish example, let's say you've observed and recorded measurements of 100 fish in Bass Lake every month for the past year. [Balık örneğimize geri dönersek, diyelim ki son bir yıldır Bass Gölü'nde her ay 100 balığı gözlemlediniz ve ölçümlerini kaydettiniz.] You want to describe these measurements and maybe create a visualization of the mean and median sizes of the fish in your sample overtime. [Bu ölçümleri açıklamak ve belki de fazla mesai numunenizdeki balıkların ortalama ve medyan boyutlarının bir görselleştirmesini oluşturmak istiyorsunuz.] This would fall into the category of descriptive statistics because you're just describing the sample. [Bu, yalnızca örneği tanımladığınız için tanımlayıcı istatistik kategorisine girer.] But you're probably not just randomly collecting data on some fish in the Lake, just for fun. [Ama muhtemelen göldeki bazı balıklarla ilgili verileri sadece eğlence olsun diye rastgele toplamıyorsunuz.] You actually would want to know if the entire population of fish in the Lake has changed significantly in size over the past year. [Aslında, göldeki tüm balık popülasyonunun son bir yılda boyut olarak önemli ölçüde değişip değişmediğini bilmek istersiniz.] Maybe you even want to predict for each month of the next year what the mean and median sizes of the fish might be. [Hatta belki de gelecek yılın her ayı için balığın ortalama ve medyan boyutlarının ne olabileceğini tahmin etmek isteyebilirsiniz.] This is when you would use inferential statistics to extrapolate or generalize from what you observe in your sample to the whole population. [Bu, örnekleminizde gözlemlediğiniz şeylerden tüm popülasyona tahminde bulunmak veya genelleme yapmak için çıkarımsal istatistikleri kullanacağınız zamandır.]