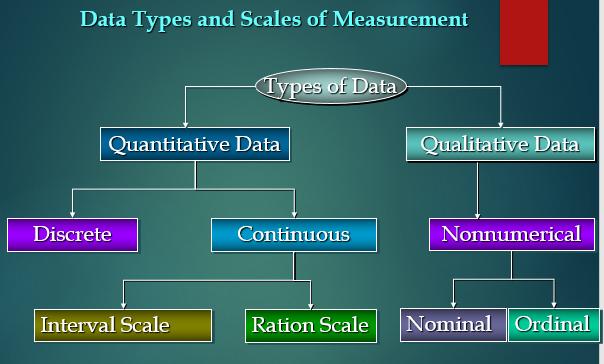
Data Types in Statistics

**Exploratory Data Analysis (EDA)**, which is one of the most important components in the pipeline of a Machine Learning Project.



There are some descriptive statistics that we can calculate for interval data such as :

* Central measures of tendency (mean, median, mode)
* Range (minimum, maximum)
* Spread (percentiles, interquartile range, and standard deviation).

**Quantitative Data**

**1.** It can be expressed as a number, so it can be quantified. In simple words, it can be measured by numerical variables such as

* “how many,
* “how much” and
* “how often”

**2.** These are easily open for statistical manipulation and can be represented by a wide variety of statistical types of graphs and charts like**line charts**, **bar graphs**, **scatter plots**, etc.

**Examples of quantitative data:**

* Scores of tests and exams e.g. 74, 67, 98, etc.
* The weight of a person.
* The temperature in a room.

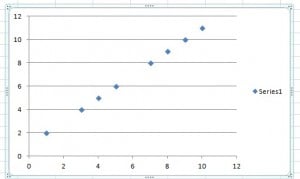
**There are 2 general types of quantitative data:**

* Discrete data
* Continuous data

### Discrete Data

## What is a Discrete Variable?

**Discrete variables**are countable in a finite amount of time. For example, you can count the change in your pocket. You can count the money in your bank account. You could also count the amount of money in everyone’s bank accounts. It might take you a long time to count that last item, but the point is—it’s still countable.

[](https://www.statisticshowto.com/wp-content/uploads/2013/09/scatter-plot-2.jpg)

*Discrete variables on a scatter plot.*

**1.**It shows the count that involves only integers and we cannot subdivide the discrete values into parts.

**For Example,** the number of students in a class is an example of discrete data since we can count whole individuals but can’t count like 2.5, 3.75, kids.

**2.** In simple words, discrete data can take only certain values and the data variables cannot be divided into smaller parts.

**3.** It has a limited number of possible values **e.g. days of the month.**

**Examples of discrete data:**

* The number of students in a class.
* The number of workers in a company.
* The number of test questions you answered correctly.

## What is a Continuous Variable?

**Continuous Variables** would (literally) take forever to count. In fact, you would get to “forever” and never finish counting them. For example, take age. You can’t count “age”.**Why not?** Because it would literally take forever. For example, you could be:  
25 years, 10 months, 2 days, 5 hours, 4 seconds, 4 milliseconds, 8 nanoseconds, 99 picosends…and so on.

[](https://www.statisticshowto.com/wp-content/uploads/2009/08/clock.jpg)

*Time is a continuous variable.*

**Qualitative Data**

**1.**Qualitative data can’t be expressed as a number, so it can’t be measured. It mainly consists of words, pictures, and symbols, but not numbers.

**2.** It is also known as **Categorical Data** as the information can be sorted by category, not by number.

**3.**These can answer the questions like:

* “how this has happened”, or
* “why this has happened”.

**Examples of qualitative data:**

* Colors e.g. the color of the sea
* Popular holiday destinations such as Switzerland, New Zealand, South Africa, etc.
* Ethnicity such as American Indian, Asian, etc.

**In general, there are 2 types of qualitative data:**

* Nominal data
* Ordinal data.

### Nominal Data

**1.**This data type is used just for labeling variables, without having any quantitative value. Here, the term ‘nominal’ comes from the Latin word **“nomen”** which means**‘name’**.

**2.**It just names a thing without applying for any particular order. The nominal data sometimes referred to as**“labels”.**

**Examples of Nominal Data:**

* Gender (Women, Men)
* Hair color (Blonde, Brown, Brunette, Red, etc.)
* Marital status (Married, Single, Widowed)

As you can observe from the examples there is no intrinsic ordering to the variables.

Eye color is a nominal variable having a few levels or categories such as Blue, Green, Brown, etc and there is no possible way to order these categories in a rank-wise manner i.e, from highest to lowest or vice-versa.

### Ordinal Data

**1.** The crucial difference from nominal types of data is that Ordinal Data shows where a number is present in a particular order.

**2.**This type of data is placed into some kind of order by their position on a scale. Ordinal data may indicate superiority.

**3.** We cannot do arithmetic operations with ordinal data because they only show the sequence.

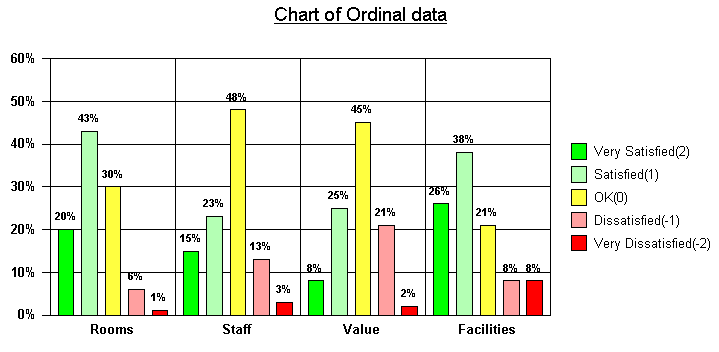
**4.**Ordinal variables are considered as “in-between” qualitative and quantitative variables.

**5.** In simple words, we can understand the ordinal data as qualitative data for which the values are ordered.

**6.** In comparison with nominal data, the second one is qualitative data for which the values cannot be placed in an order.

**7.** Based on the relative position, we can also assign numbers to ordinal data. But we cannot do math with those numbers.**For example,** “first, second, third…etc.”

**Examples of Ordinal Data:**

* Ranking of users in a competition: The first, second, and third, etc.
* Rating of a product taken by the company on a scale of 1-10.
* Economic status: low, medium, and high.
* Week Days
* Month Names

**Interval Data**

**1.** These data types are measurable and ordered with the nearest items but have no meaningful zero.

**Let’s understand the meaning of “Interval Scale”:**

In the Interval scale, the term ‘Interval’ signifies space in between, which is a significant thing to recall as interval scales not only educate us about the order but in addition, give information about the value between every item.

**2.**Fundamentally, we can show interval data in the same way as ratio data, but the thing that we have to note is their characterized zero points.

**3.**Hence, with the help of interval data, we can easily correlate the degrees of the data and also add or subtract the values.

**4.**There are some descriptive statistics that we can calculate for interval data such as :

* Central measures of tendency (mean, median, mode)
* Range (minimum, maximum)
* Spread (percentiles, interquartile range, and standard deviation).

These are not the only statistical things to be calculated, but we can calculate more things also.

**Examples of Interval data:**

* Temperature (°C or F, but not Kelvin)
* Dates (1055, 1297, 1976, etc.)
* Time Gap on a 12-hour clock (6 am, 6 pm)

**Ratio Data**

**1.** These data are also in the ordered units that have the same difference.

**2.** Ratio values are the same as interval values, but the only difference is that Ratio data do have an absolute zero. **For Example,**height, weight, length, etc.

**3.**These are measured and ordered with equidistant items with a meaningful zero and never be negative like interval data.

Let’s understand this with an outstanding example- **Measurement of heights**.

Height can be measured in units like centimeters, inches, meters, or feet and it is not possible to have a negative value of height.

**4.**It enlightens us regarding the order for variables, the contrasts among them, and they have absolutely zero.

**5.**Ratio data is fundamentally the same as interval data, aside from zero means none.

**6.**The descriptive statistics which we can calculate for ratio data are the same as interval data such as :

* Central measures of tendency (mean, median, mode)
* Range (minimum, maximum)
* Spread (percentiles, interquartile range, and standard deviation).

**Example of Ratio data:**

* Age (from 0 years to 100+)
* Temperature (in Kelvin, but not °C or F)
* Time interval (measured with a stop-watch or similar)

For the above examples of ratio data, we see that there is an actual and meaningful zero-point like the age of a person, absolute zero, distance calculated from a specified point or time all have real zeros.

--------------------------------------------------------------------------

Ref: <https://www.analyticsvidhya.com/blog/2021/06/complete-guide-to-data-types-in-statistics-for-data-science/>

Ref: https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/discrete-vs-continuous-variables/