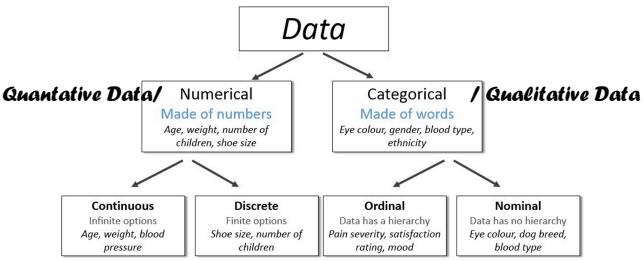
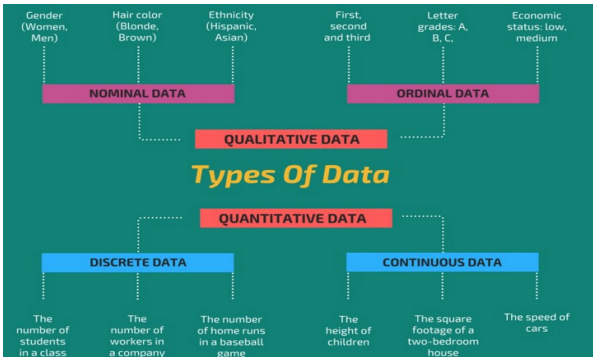
Data Types

24 June 2021 08:39





Quantitative Data / Numerical Data -

1. easiest to explain. It tries to find the answers to questions such as
   - "how many,
   - "how much" and
   - "how often"

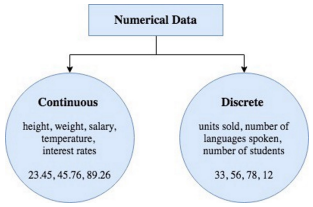
   Any Data where data points are exact Numbers
2. It can be expressed as a number, so it can be quantified. (it can be measured by numerical variables).
3. These are easily open for statistical manipulation and can be represented by a wide variety of statistical types of graphs and charts like **line charts**, **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:

a. Discrete data -
b. Continuous data -



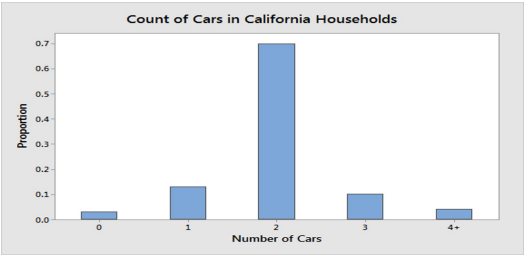
---

1. Discrete Data -

- 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.
- Discrete data can take only certain values and the data variables cannot be divided into smaller parts.
- It has a limited number of possible values **e.g. days of the month.**

Examples of discrete data:

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



Qualitative Data/ Categorical Data -

1. These can answer the questions like:
   - "how this has happened", or
   - "why this has happened".

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

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:

1. Nominal data -
2. Ordinal data -

---

1. Nominal Data -

- This data type is used just for labeling variables, without having any quantitative value. term 'nominal' comes from the Latin word **"nomen"** which means '**name**'.
- It just names a thing without applying for any particular order. The nominal data sometimes referred to as **"labels".**

Examples of Nominal Data:

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

Note - Nominal data that has no order.
if we change the order of its values, the meaning would not change.

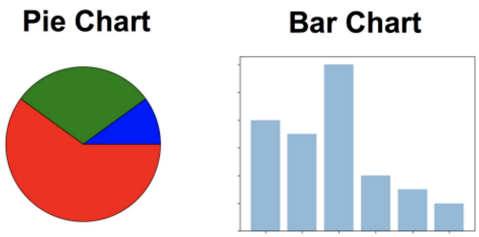
example -

When you are dealing with nominal data, you collect information through :

- **Frequencies** : The Frequency is the rate at which something occurs over a period of time or within a dataset.
- **Proportion** : You can easily calculate the proportion by dividing the frequency by the total number of events. (e.g how often something happened divided by how often it could happen)
- **Percentage**

Visualization Methods: To visualize nominal data you can use a pie chart or a bar chart.

Visualization Methods: To visualize nominal data you can use a pie chart or a bar chart.

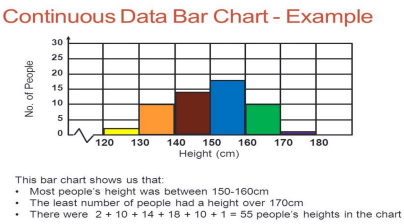


1. Continuous Data -

- It represents the information that could be meaningfully divided into its finer levels. It can be measured on a scale or continuum and can have almost any numeric value.
- **For Example,** We can measure our height at very precise scales in different units such as meters, centimeters, millimeters, etc.
- The key difference between continuous and discrete types of data is that in the former, we can record continuous data at so many different measurements such as width, temperature, time, etc.
- The continuous variables can take any value between two numbers. **For Example,** between the range of 60 and 82 inches, there are millions of possible heights such as 62.04762 inches, 79.948376 inches, etc.
- A good great rule for defining if data is continuous or discrete is that if the point of measurement can be reduced in half and still make sense, the data is continuous.

Examples of continuous data:

a. The amount of time required to complete a project.
b. The height of children.
c. The speed of cars.



In Continuous data and there are 2 types

i. Interval Data-

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

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

Temperature?
- -10
- -5
- 0
- +5
- +10

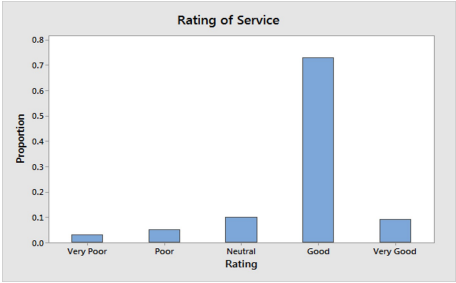
2. Ordinal Data -

- Ordinal Data shows where a number is present in a particular order.
- Ordinal data mixes of both numerical and categorical data
- This type of data is placed into some kind of order by their position on a scale. Ordinal data may indicate superiority.
- We cannot do arithmetic operations with ordinal data because they only show the sequence.
- Ordinal variables are considered as "in-between" qualitative and quantitative variables.
- In comparison with nominal data, the second one is qualitative data for which the values cannot be placed in an order.
- 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:

a. Ranking of users in a competition: The first, second, and third, etc.
b. Rating of a product taken by the company on a scale of 1-10.
c. Economic status: low, medium, and high.
d. rating a restaurant on a scale from 0 (lowest) to 4 (highest) stars

j



- ○ -5
- ○ 0
- ○ + 5
- ○ + 10
- ○ + 15

ii. Ratio Data

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

b. 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.

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

understanding 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.
- and they have absolutely zero.
- Ratio data is fundamentally the same as interval data, aside from zero means none.
- 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.

Length (inch)?
- ○ 0
- ◉ 5
- ○ 10
- ○ 15

NOTE:

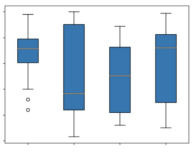
If we picked the zero-point of the scale subjectively, then at that point the data can't be ratio data and should be interval data.

Visualization Methods :

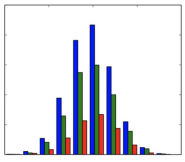
- We can use a histogram or a box-plot to visualize continuous data.
- With a histogram, you can check the central tendency, variability, modality, and kurtosis of a distribution.

Note - that a histogram can't show you if you have any outliers. This is why we also use box-plots.

Boxplot



Histogram

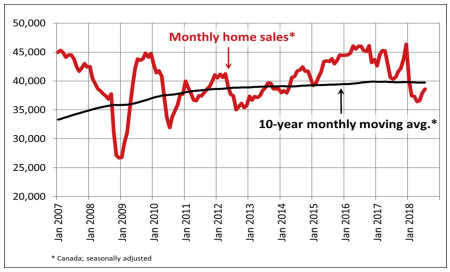


Time Series Data

- ➢ Time series data is a sequence of numbers collected at regular intervals over some period of time.
- ➢ important, especially in particular fields like finance.
- ➢ Time series data has a temporal value attached to it, so this would be something like a date or a timestamp that you can look for trends in time.

example -

- ○ we might measure the average number of home sales for many years



Text -

- ➢ Text data is basically just words.
- ➢ A lot of the time the first thing that you do with text is you turn it into numbers using some interesting functions like the bag of words formulation.