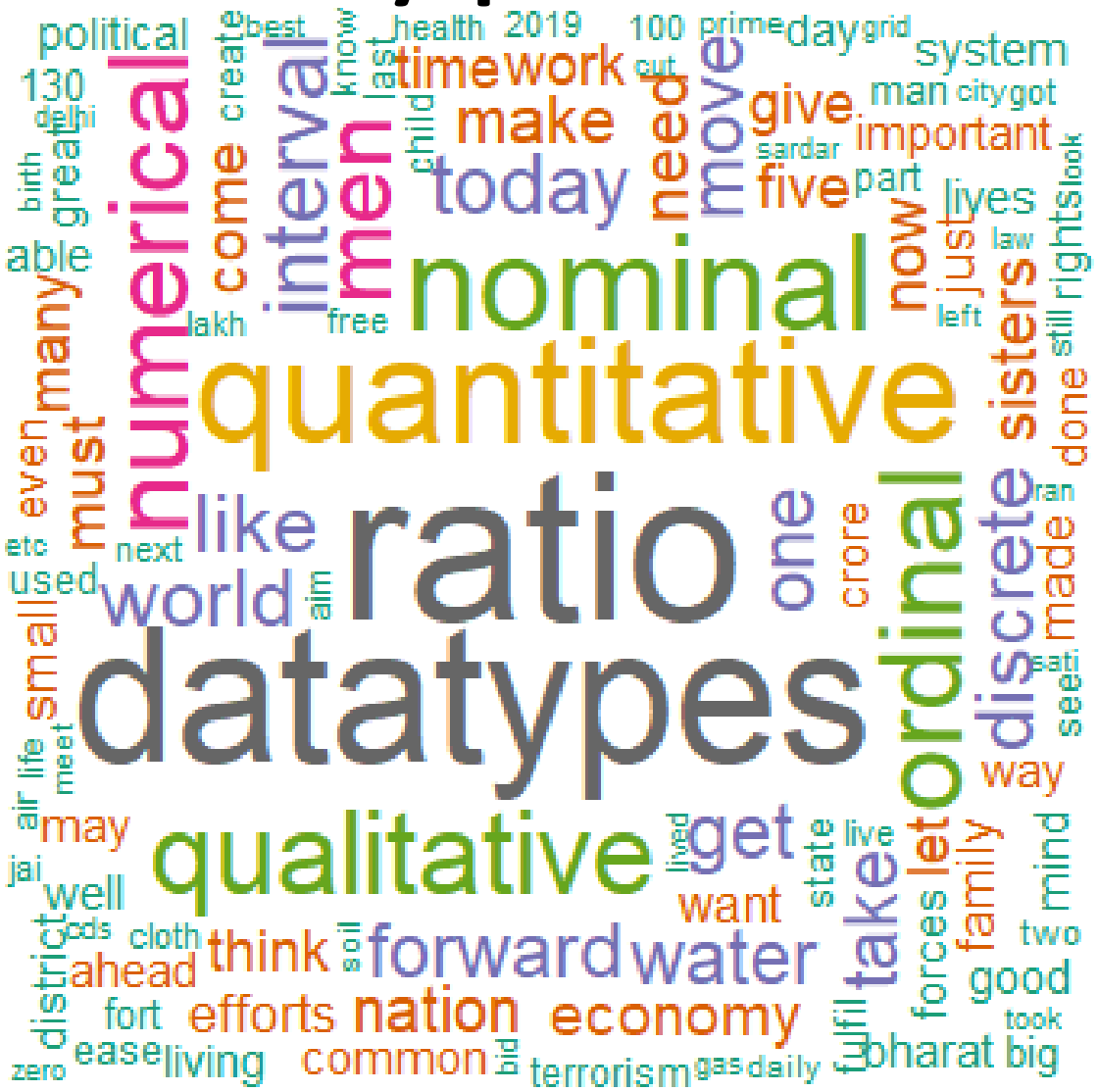
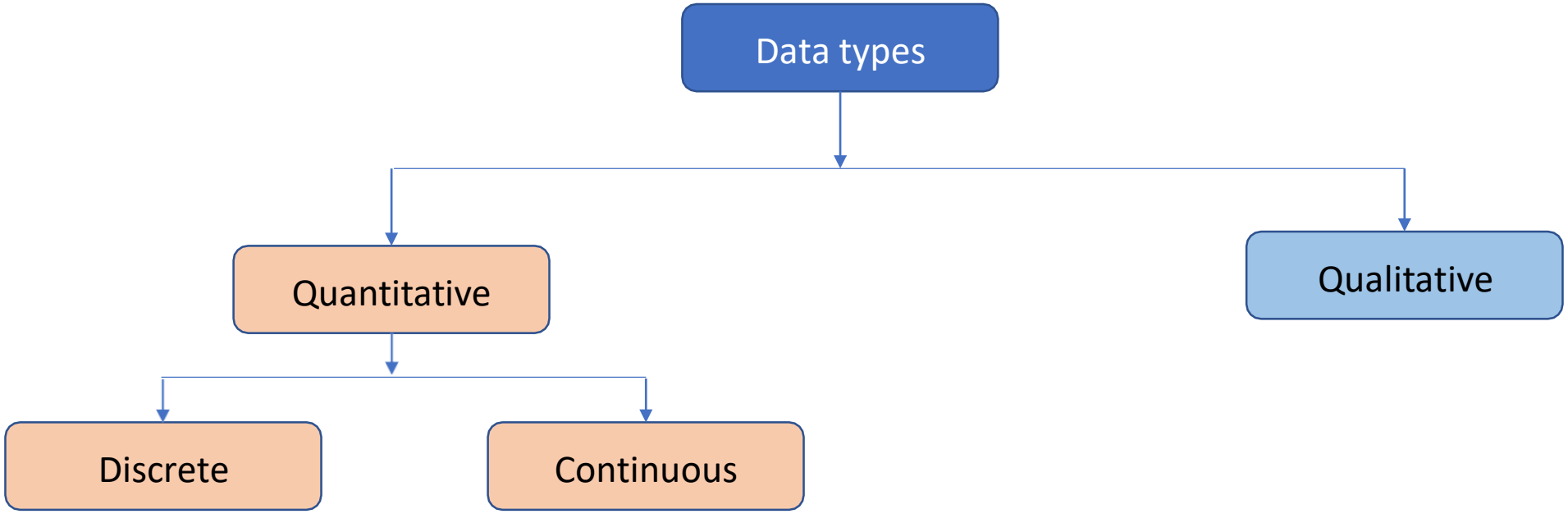
Four Types of Data



Understanding data type is an important concept in statistics, when you are designing an experiment, you want to know what type of data you are dealing with, that will decide what type of statistical analysis, visualizations and prediction algorithms could be used. Also, you can use a particular statistical measurement only for specific data types.



Qualitative & Quantitative are two basic types of data. As the name suggests, qualitative deals with the quality (Characteristics) & some statisticians calls it categorical. Quantitative deals with numbers. Below is the difference between Data Types.

Qualitative	Quantitative
Non numerical, Categories, Attributes	Numerical
Examples: Colour, Ethnicity, Gender, political leanings, religion, Education etc.	Examples: time, count, weight, length, wages, Income
Even though some data are number mathematical operations are meaningless. Zipcode and phone numbers	Mathematical operations are meaningful.

Table1: Difference between types of data

- Quantitative are of two types viz. Discrete and Continuous. They are explained below:
  - Discrete: Discrete data are countable or Finite. Finite means there are certain number of values you can pick from and countable means you can count them. This type of data can't be measured but it can be counted. These are natural numbers and are count of something.
    - Let me give you an example of countable, if you are asked to count number of cars on road during certain time of the day, it can take numbers like 100,125 or 1000 but never 100.5 or 125.72 etc.
    - Likewise, example of finite is, the number which comes on rolling a dice. There are only 6 possible choices like 1,2,3…6 but never more than 6 or 4.5 etc. likewise if you flip a coin it has only heads or tails, so, there are certain number of values you can pick from.
  - Continuous: Continuous data are uncountable or infinite or to put it differently, there are infinite number of possible values and is not countable. Usually is a measurement of something and cannot be counted. Continuous can take absolutely any value.
    - For example, a person's height or weight. Height can be 5.23, 5.24, 5.76 etc. similarly weight can be any value like 75.82 Kgs or 62.35 kgs etc. Height, weight, length, speeds, temperatures etc. are examples of continuous data.

Levels of measurement/data

In the figure above as we move from left to right, we see an increase in order of information i.e. Ordinal will have the characteristics of Nominal and something more, similarly, interval will have the characteristics of Ordinal and something more and Ratios will have characteristics of Interval and something more… Let's see what each one of these data levels means.

- Nominal:
  - Nominal means a name only, Nominal scales are used for labelling variables, without any quantitative value, it is a categorical data which has no order. Red car, Blue car, yellow car, Black Car, White car etc. are just categories, they don't have any order to them. All you can do is list them out. For example, Flavour, Religion, Ethnicity, Gender etc. these are nominal data without any order. For statistical analysis we can assign numbers to these categories for example: Red=1, White=2 and Black=3, these numerical values assigned does not have any mathematical significance. A sub-type of nominal scale with only two categories (e.g. male/female, hot/cold, good/bad) is called "dichotomous."
- Ordinal:
  - it is a categorical data which has an implied order. Like size of clothing as Small, Medium and Large or Likert scale questions having a scale from 1 to 5 for example Agree=1, neutral=2 disagree=3 etc. The numbers do not have any mathematical significance, but they are the labels. Subtraction and division don't make any mathematical significance. For example, Ranking (ranking in school exams i.e 1^{st},2^{nd} and 3^{rd} etc., ranking in the Army i.e Captain, Major, Colonel etc.), survey done on Likert scale, educational background(under graduate, Postgraduate, Doctoral etc.).in all the above examples there is an order associated with it.

- Interval: Interval values are Categorical and ordered data in addition to that they have scale to them. Interval values data don’t have a true zero. True zero means absence of the variable. For example, zero degrees temperature does not mean there is no temperature, it just means, it is too cold. Here Addition and subtraction are significant, but division and multiplication are not. With interval data, we can add and subtract, but we cannot multiply, divide or calculate ratios. Good examples are Time, temperature
- Ratio: Ratio values are Categorical, ordered data having scale to them and they have a natural or true zero. They can be discrete or continuous. Good examples are height, weight, length, Duration etc. Since they have natural zero it allows for a wide range of both descriptive and inferential statistics to be applied. For example, zero dollars means there is no money. If length of two rods is 2 fts and 4 fts…then we can say that one rod is twice the length of the other. Sales figures, Sales of zero means that you sold nothing and so sales didn’t exist.

Summary

- In this post, you discovered the different data types that are used by Data Scientists. You learned the difference between discrete & continuous data and learned about levels of data viz. nominal, ordinal, interval and ratio measurement scales.
  - Nominal are used to “name,” or label a data.
  - Ordinal provides information about the order of values,
  - Interval give us the order of values plus the ability to quantify the difference between each one.
  - Finally, Ratio give us the order, interval values, plus the ability to calculate ratios since a “true zero” can be defined.