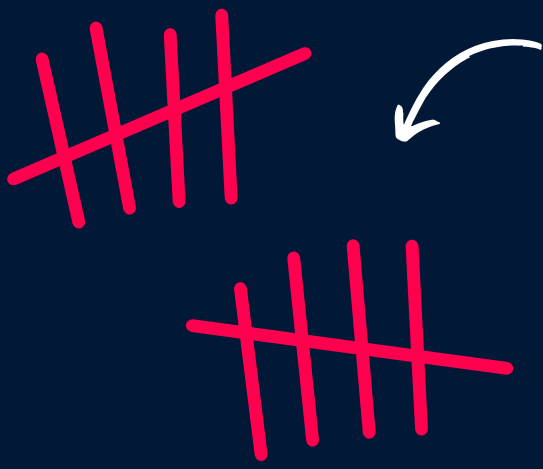
DATA SCIENCE INFINITY

4 TYPES OF DATA

1

2



3

4

AN OVERVIEW

What?

Data is everywhere! This is even more true for those in Data Science & Analytics - it is the lifeblood of all things we do!

There are several different *types* of data - and knowing the difference is important to ensure we can **process it** in the right way, and **get the results** we need!

At a high level, we can break data down into two main types; **Numeric** & **Categorical** (also sometimes called Quantitative & Qualitative)

Numeric

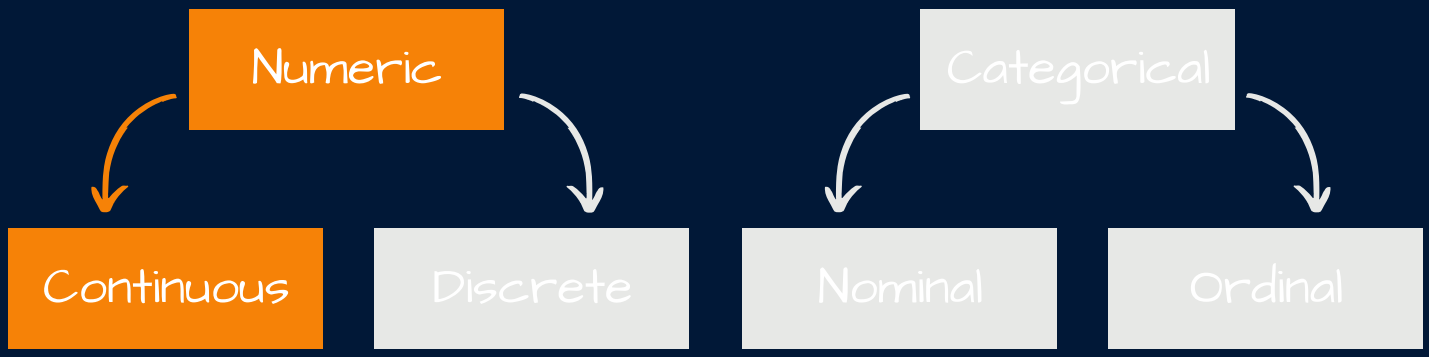
Categorical

The simplest possible way to differentiate these two categories is that **numeric data can be expressed as a number, while categorical data cannot**.

These two categories of data can be broken down further - and this will all start to make more sense - let's take a look!

DATA SCIENCE INFINITY

Continuous data...



Continuous data is numeric, but exists in **fractional** form - in that it represents information that can be divided to a more granular level.

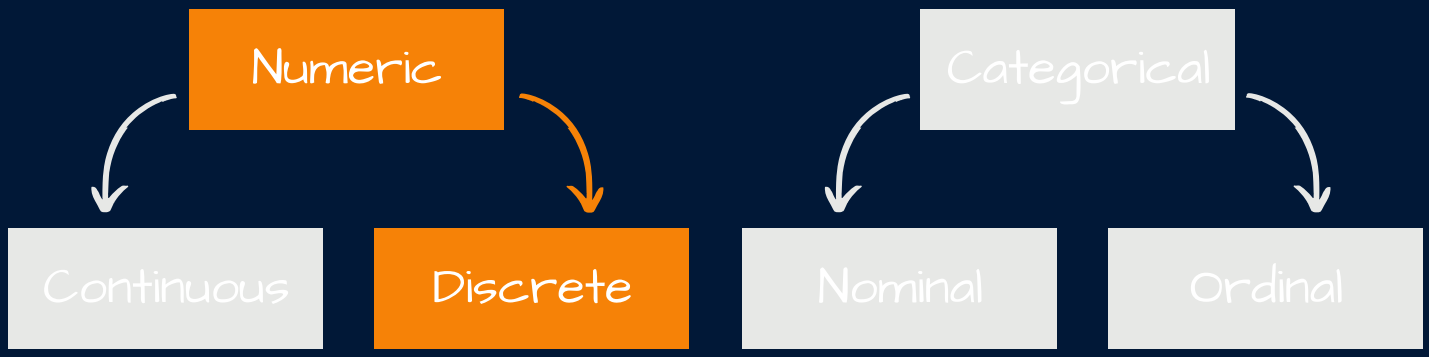
For example, the NBA Player **LeBron James** is listed as being 2.06m tall....

...but if we had a high-tech distance measuring device, we might find that he is actually 2.064759m tall.

Continuous data has an infinite number of possible values within.

Other examples could be the speed of a truck, the time taken to run a race, or the size of a room.

Discrete data...



Like continuous data, **discrete data** is *also* numeric, but exists in **whole** form - in that it represents information that is countable and cannot be divided into smaller forms.

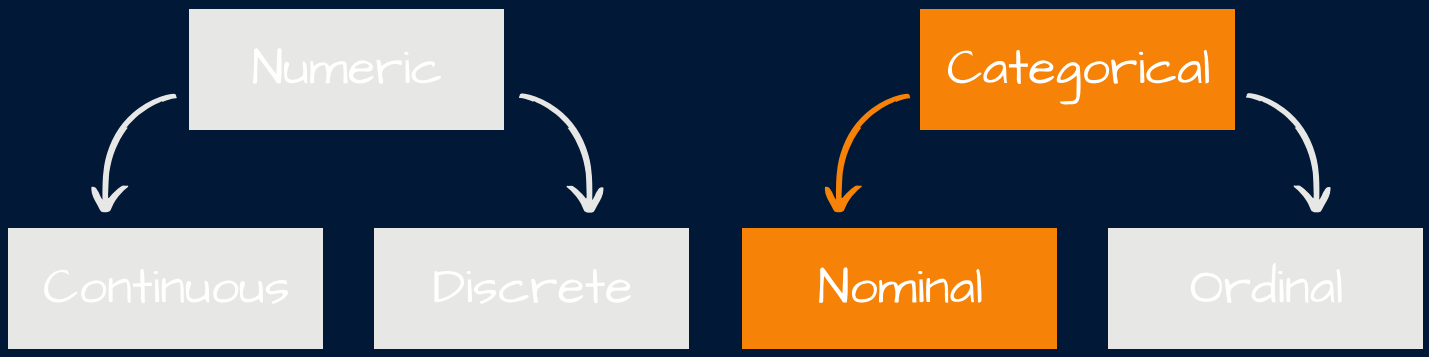
For example, the football player **Cristiano Ronaldo** has scored 38 goals this season...

...unlike LeBron James' height, this value can't be broken down further - there are no half or quarter goals.

Other examples could be the number of students in a school, the number of entrants in a competition, or the number of days in a year...

DATA SCIENCE INFINITY

Nominal data...



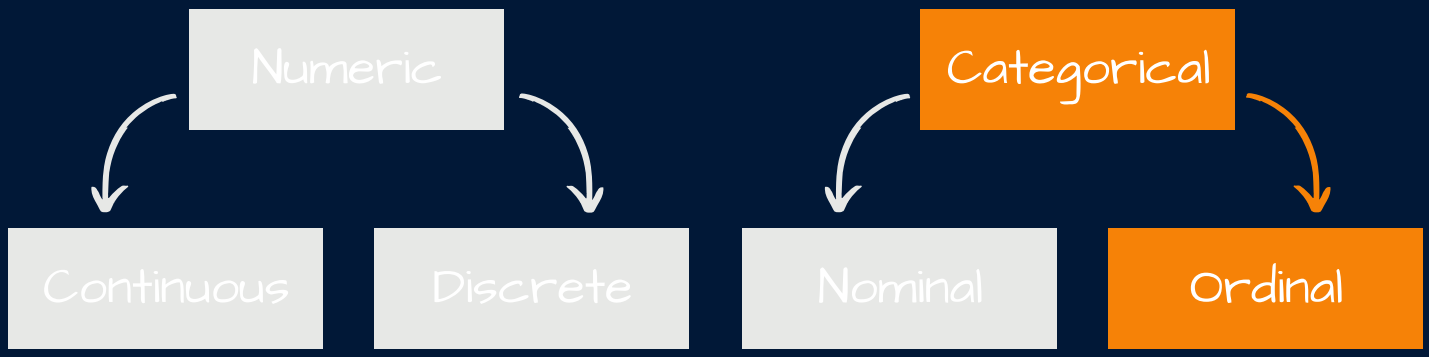
Nominal data is a type of categorical data - unique in the fact that the data-points **do not have an order**. Nominal comes from the Latin "nomen" which simply means "name"

For example, the social media brands to the right; Instagram, TikTok, Twitter & Facebook...

On their own, they are just a list of **names, labels or categories** - without any one of them being greater or smaller, or better or worse than the others.

Other examples could be; people's first names, countries, or colours

Ordinal data...



Like nominal data, **ordinal data** is a type of categorical data. The difference is that ordinal data **does contain an underlaying order or ranking**

For example, t-shirt sizes where a Medium is bigger than a Small, a Large is bigger than a Medium (and much bigger than a Small)

Even if these ordinal values take on numeric form (i.e. 1st, 2nd, 3rd) we avoid applying arithmetic to them because they generally only show the sequence, not the scale.

Other examples could be; Good, Average, Bad values for customer satisfaction, or letter grades for an exam result of A, B, C etc...

Why does this matter?

While these data types might seem obvious to some, they are important and underpin a lot of other concepts in Statistics & in Data Science!

Knowing the difference is important to ensure we can **process it** in the right way, and **get the results** we need!

We may need to know which type of **Hypothesis Test** to apply to measure a change or difference, and knowing the type of data we're dealing with is crucial.

We may need to clean or pre-process our data for **Machine Learning** - and knowing the different data types will determine how we do this in an appropriate manner.

There are many other areas where these simple concepts help us work efficiently & effectively so it's always good to keep them in mind!

Happy Learning!