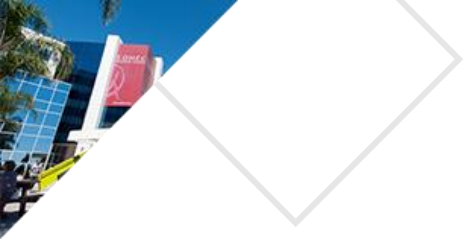


TYPES OF DATA

TYPES OF DATA

- In **programming** we speak about data types.
- In **statistics** we speak about types of data (of type of variable).
- They are not the same !

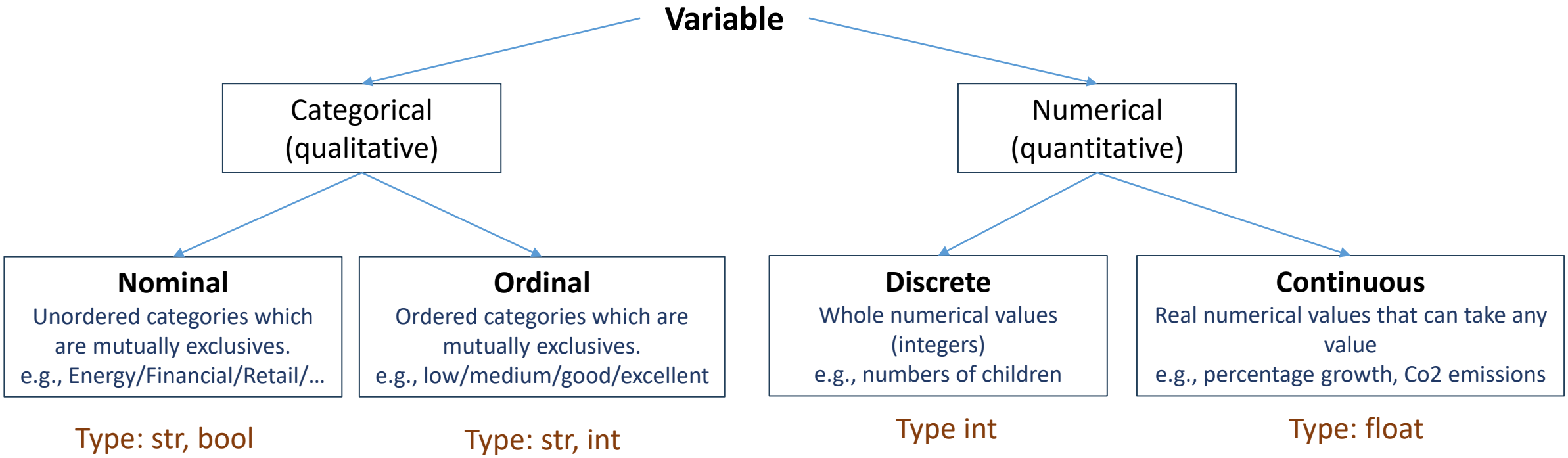
- Let's give some definitions.

DEFINITIONS

- A **variable** is a characteristic, number or quantity that can be measured (or counted).
- Each of this measurements is called a **data point**.
- We can measure several variables at the same time. In this case we say that out data is multidimensional.
- We can write our data as a table. Each column corresponds to one variable, and each line corresponds to a simultaneous measurement (a **record**, or by extension, a **data point**).
- Such a table, is a very simple example of **database**, which is an organised collection of data and information.

TYPE OF DATA

Variables are usually classified in several groups.



Approximate correspondances between data types and types of data

TYPE OF DATA DETERMINES WORFLOW AND ANALYSIS

It determines for instance:

- Type of visualisation;
- Type of statistical analysis that can be used;
- Type of Machine learning model used in AI;
- Type of data types used in a program to treat that data;
- Even the type of database used to store the data.

WHAT DATA LOOKS LIKE

Variables

*(also called **features, factors, attributes, predictors, covariates**, and many other names)*

Records or Data points

id_invoice	amount	age customer	items bought	credit card	Satisfaction_level
1	163.5	43	7	0	good
2	138.8	39	5	0	very good
3	175.9	48	7	1	low
4	157.5	45	11	1	low
5	600	43	2	0	very good
6	132.8	32	5	0	good
7	165.7	42	3	1	very good
8	134.1	39	2	1	good
9	174	39	1	1	very good
10	183.7	41	11	1	medium
11	157.8	39	17	1	medium
12	157.4	37	11	1	very good
13	154.3	43	16	1	low
14	169.2	37	3	1	very good
15	162.8	44	12	1	very good
16	108.4	38	6	1	good
Type of data	Categorical	Continuous	Discrete/Continuous	Discrete	Categorical	Ordinal
Possible data type	Int, str	float	Int, float	int	Bool, int	str, int

VISUALISING VARIABLE DATA

You need to tabulate data

Type data / Type chart	Pie Chart	Line Chart	Bar Chart	Histogram	Scatter Plot (2 variables or more)
Categorical	X		X		
Ordinal	X		X		
Discrete		X	X	X	
Continuous (1 variable)		X		X	
Continuous (2 variables)				X	X

This table just give some ideas on the type of graphs that are usually more useful for a given type for data.
But it will always depend on the context, and *the question to ask our data*.