

# Data Mining 4315: Lecture 01 (Page 1)

J Basravi - Jan 26, 2023 6:30pm

A **Data Set** is a collection of **data objects**.

Objects are also sometimes called:

**samples** (in statistics)  
**examples or instances** (in machine learning)  
**tuples** (in databases)

Objects are described by **attributes**.

Attributes are also called:

**dimensions**  
**features**  
**variables**

**OBJECTS (ROWS)**

**DATA SET**

| Name   | FName | City | Age | Salary |
|--------|-------|------|-----|--------|
| Smith  | John  | 3    | 35  | \$280  |
| Doe    | Jane  | 1    | 28  | \$325  |
| Brown  | Scott | 3    | 41  | \$265  |
| Howard | Shemp | 4    | 48  | \$359  |
| Taylor | Tom   | 2    | 22  | \$250  |

**ATTRIBUTES (COLUMNS)**

**Numeric** attributes have **number values**.

For example, *age* = 6, *weight* = 155, *temp* = 62.2

**Non-numeric** attributes are **qualitative (not number values)**.

For example, *color* = yellow, *name* = Michael, *hungry* = yes

**Interval-scaled**

attributes are numeric and have values on a numerical scale with equal intervals between values. There is **NO ZERO POINT** defined.

For example, *IQ scores* fall on a numerical IQ scale and describe intelligence. There is no value on the scale (not even IQ = 0) that indicates a total absence of intelligence, so it is interval-scaled.

**Ratio-scaled** attributes are numeric and have values on a numerical scale with equal intervals between values. There **IS A ZERO POINT** defined.

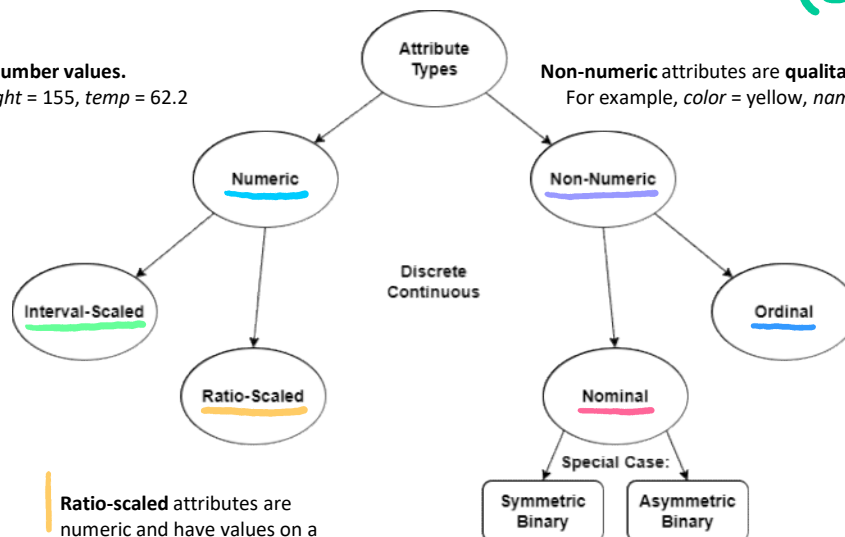
For example, a car's *speed* falls on a numerical scale and describes its rate of travel. There is a value on this scale that describes the total absence of speed (0 mph), so it is ratio-scaled.

A **discrete** attribute has a **finite** number of values.

For example, *color* (there is a finite number of colors).

A **continuous** attribute has an **infinite** number of values.

For example, *height* (there is an infinite number of heights).



**Ordinal** attributes are non-numeric and describe a variable with values that have a **meaningful order**.

For example, *levels of agreement* (strongly agree, somewhat agree, neutral, somewhat disagree, strongly disagree).

**Nominal** attributes are non-numeric and describe a variable with values where **order does NOT matter**.

For example, *eye color* (blue, green, brown, etc.)

There is a special kind of nominal attribute called a **binary attribute**, which can **only have two values**.

For example, *sex* = male/female, *likes chocolate* = yes/no

A **symmetric binary attribute** is a binary attribute where each of its two values have **equal importance** (*wears glasses* = yes/no, *has a pet* = yes/no).

An **asymmetric binary attribute** is a binary attribute where, of its two values, **one is more important** (*medical test* = positive/negative, *criminal record* = yes/no).