

# Understanding Data in Computing

---

## Introduction

In the world of computing, **data** is essentially raw, unorganized facts and figures. It can be a collection of numbers, words, images, or symbols that a computer can process. On its own, data may not have much meaning, but when it's processed and organized, it becomes valuable **information**.

## Types of Data: Structured vs. Unstructured

### Structured Data

Structured data is highly organized, easily searchable, and follows a predefined schema (like tables with rows and columns).

#### Key Characteristics:

- **Predefined Schema:** Each field has a fixed format (e.g., Date field only has dates).
- **Quantitative:** Consists of numbers and measurable values.
- **Easy to Query:** Searchable using SQL.
- **Storage:** Stored in relational databases and data warehouses.

#### Examples:

- Excel spreadsheets with customer records.
- Employee databases (ID, name, salary).
- Point-of-sale transaction records.

### Unstructured Data

Unstructured data does not have a predefined schema, making it more complex to analyze.

#### Key Characteristics:

- **No Predefined Schema.**
- **Qualitative.**
- **Difficult to Query:** Requires NLP, image recognition, etc.
- **Storage:** Stored in data lakes or NoSQL databases.

#### Examples:

- **Text:** Emails, PDFs, social media posts.
- **Multimedia:** Images, audio, video.
- **Sensor Data:** IoT devices, weather sensors.

## Comparison Table

Feature	Structured Data	Unstructured Data
Organization	Fixed schema (tables)	No predefined format
Type	Quantitative (numbers, dates)	Qualitative (text, media)
Ease of Analysis	Easy with SQL	Complex, needs advanced tools
Storage	Relational DBs	Data Lakes, NoSQL
Flexibility	Less flexible	Highly flexible

## Quantitative vs. Qualitative Data

### Quantitative Data

- Deals with numbers and measurable facts.
- Analysis: Mathematical/statistical methods.
- Collection: Surveys, experiments, web analytics.

**Examples:** Temperature (25°C), Height (175cm), Number of students (30).

### Qualitative Data

- Descriptive, conceptual, non-numerical.
- Analysis: Thematic, interpretative.
- Collection: Interviews, observations, open-ended surveys.

**Examples:** Car color ("blue"), feelings ("happy"), interview transcripts.

## Comparison Table

Feature	Quantitative Data	Qualitative Data
Type	Numbers, measurable	Descriptions, characteristics
Asks	How many? How much? How often?	Why? How?
Analysis	Statistical, graphs, charts	Thematic, interpretation
Form	Numbers, tables	Text, narratives, images
Objectivity	Objective	Subjective

## Numerical Variables

### Continuous Variables

- Can take infinite values in a range.
- Measured with precision (decimals possible).

**Examples:** Height (175.2cm), Temperature (25.35°C), Time (11.52s).

## Discrete Variables

- Countable, specific values (no fractions).

**Examples:** Number of students, dice roll, cars in a parking lot.

## Summary Table

Feature	Continuous Variable	Discrete Variable
Values	Any value in range	Specific whole numbers
Nature	Measurable	Countable
Examples	Height, weight, time	Dice rolls, class size

## Qualitative (Categorical) Variables

### Nominal Variables

**Examples:** Eye color, country, types of pets.

### Ordinal Variables

**Examples:** Satisfaction ratings, education levels, t-shirt sizes.

### Binary Variables

**Examples:** Yes/No, True/False, Pass/Fail.

## Summary Table

Feature	Nominal	Ordinal	Binary
Meaning	Categories, no order	Ordered categories	Two categories only
Example	Eye color	Education level	Pass/Fail

## Measures of Central Tendency

### Mean

Average of all values. Sensitive to outliers.

### Median

Middle value in sorted dataset. Resistant to outliers.

### Mode

Most frequent value. Works with categorical data.

## Measures of Spread

- **Range:** Max - Min
- **Variance:** Average squared deviation from mean
- **Standard Deviation:** Typical distance from mean
- **Interquartile Range (IQR):** Spread of middle 50%

## Shape of Data Distribution

### Skewness

- Symmetrical: Mean = Median = Mode
- Right Skew: Mean  $>$  Median
- Left Skew: Mean  $<$  Median

### Common Distributions

- Normal (bell curve).
- Uniform (all values equally likely).
- Skewed (left or right).
- Bimodal (two peaks).