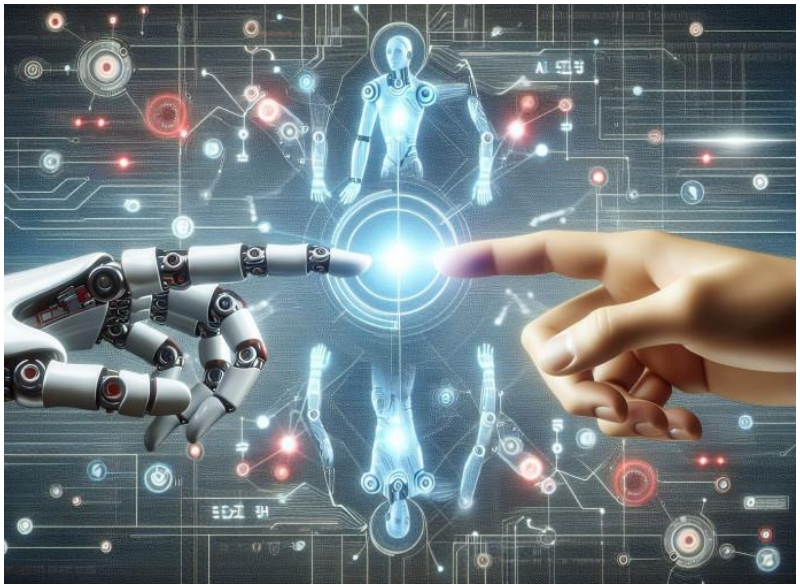
# Lecture 2
# Knowledge Representation
# (1- Data and Information)

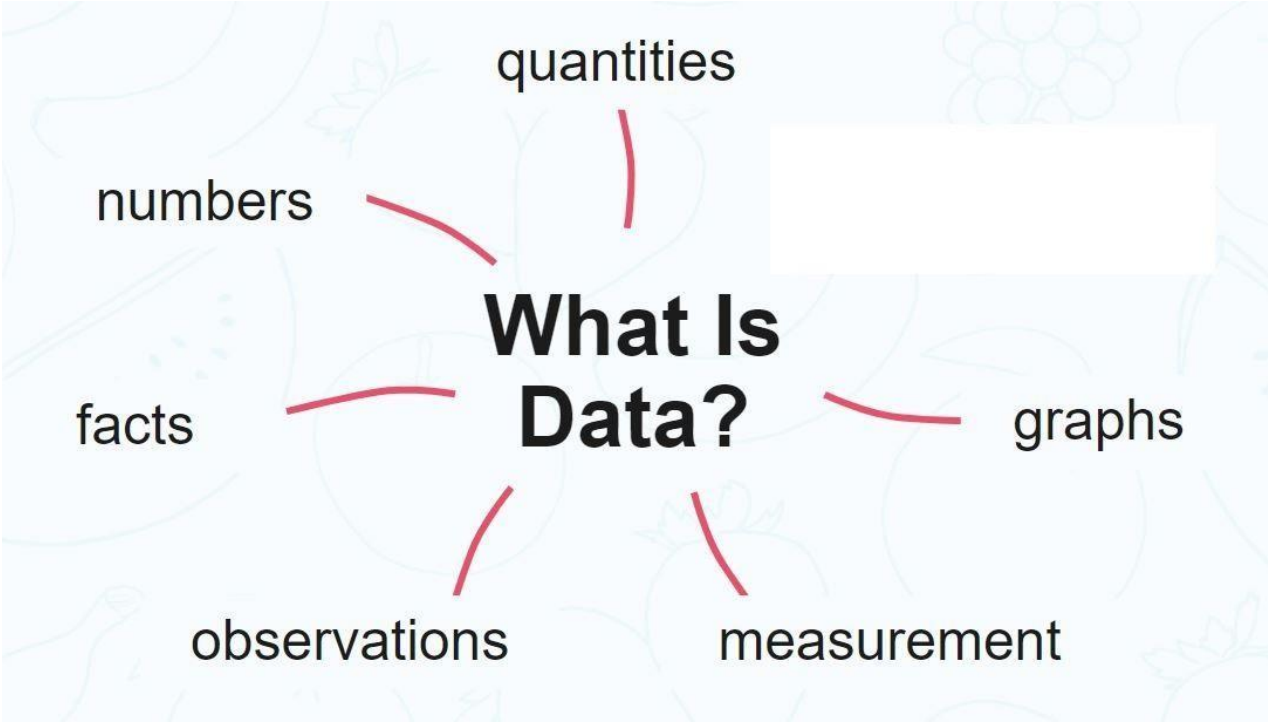


**Dr. Fatma Eskander**
**Math.department, Faculty of Science, Mansoura University**

# Data

# Data

- **Data** are individual facts, observations, statistics, characters, symbols, images, numbers, and more that are out of context, have no meaning, and are difficult to understand. They are often referred to as **raw data**.



# Types of Data

Research data are generally classified either as **quantitative** or **qualitative**.

**Quantitative data:**

These are data which can be counted or expressed in numerical values.

e.g. nb of students, age, grades, test scores…

**Qualitative data:**

These are descriptive data which has no numerical values.

e.g. A person's attitude, perception, or feelings

# Types of Data

Based on their sources, they fall under two categories:

- **Primary data;**
- **Secondary data**

**Primary data (**first-hand information**)**

Primary data are originated by a researcher for the specific purpose of addressing the problem at hand. They are collected directly from the subjects being studied.

**Secondary data:**

These are data already available. They have been collected from other available sources.

# Types of Data

Secondary data may be published or unpublished.

**Secondary data include:**

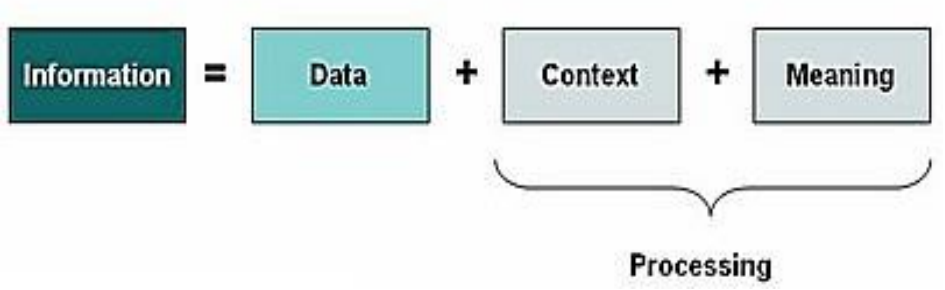
- Files/records
- Computer data bases
- Industry or government reports
-Documents (budgets, organizational charts, policies and procedures, maps, monitoring reports)
- newspapers and television reports

# Information

# Information

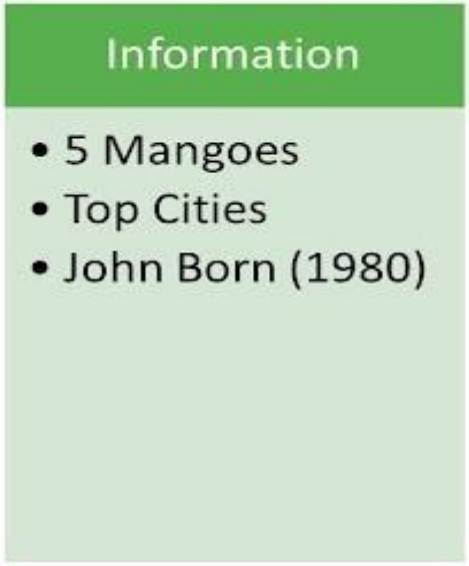
- **Data** is meaningless in itself, but once processed and interpreted, it becomes **information** which is filled with meaning.

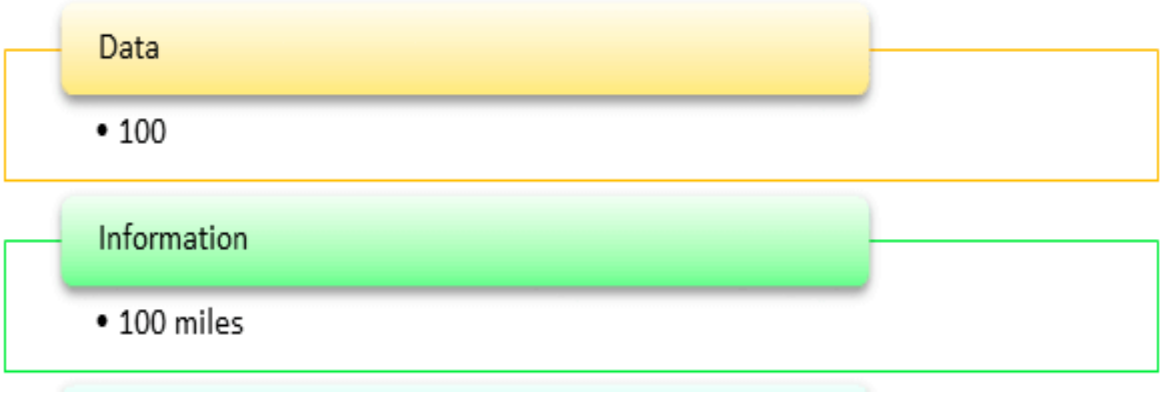
- **Information** is a set of data that is processed in a meaningful way according to the given requirement. It is processed, structured, or presented in a given context to make it meaningful and useful.



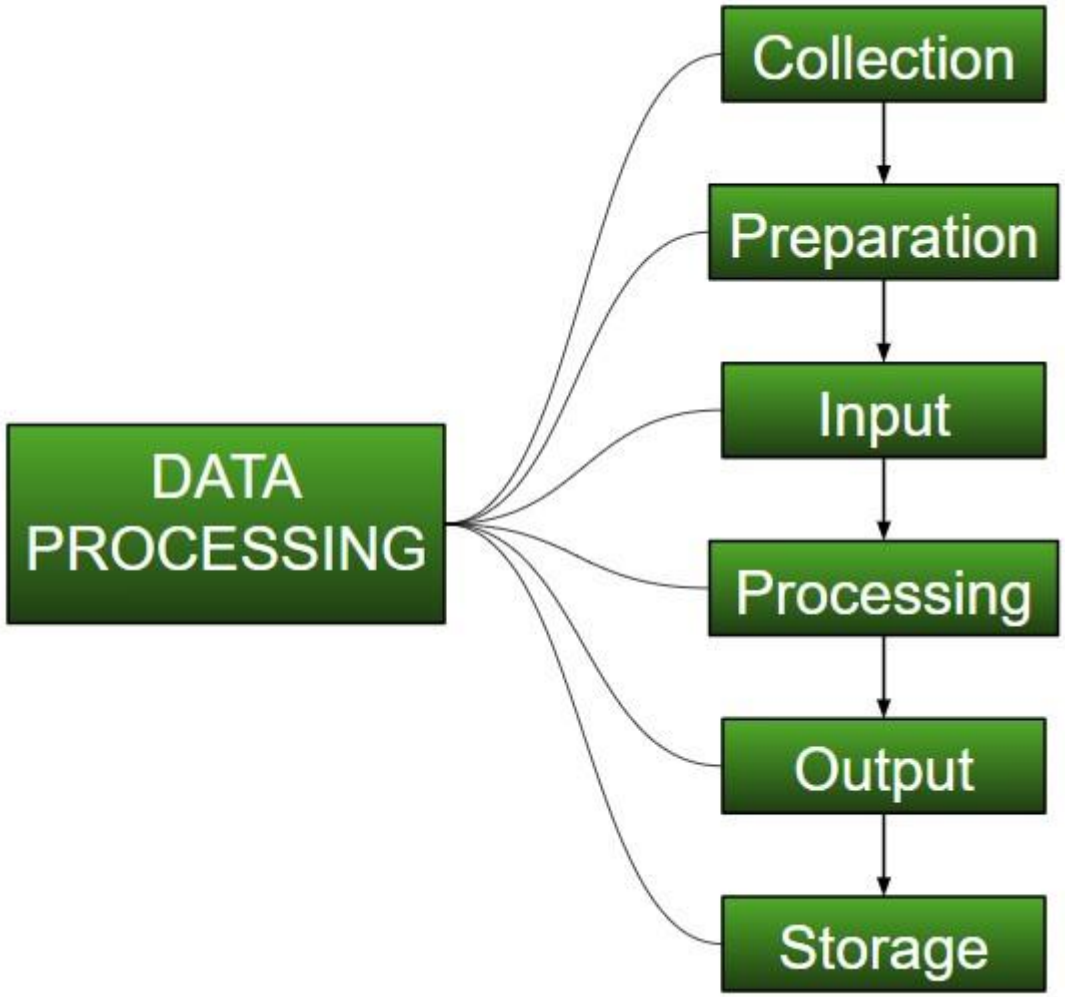
# Information

Following two figures gives an idea of data and information

# Data Processing



# Data Processing

Data in its raw form is not useful to any organization.

**Data processing** is the method of collecting **raw data** and translating it into **usable information**.

**The Data Processing Operations**

## A. Data Collection

**Methods of data collections**

**1. Observation** **2. Questionnaire** **3. Interview**

**4. Surveys** **5. Expérimental devices**

# Data Processing

## B. Data preparation (Validation) :

Ensuring that supplied data is correct and relevant.

A process of examining the collected raw data to **detect errors and correct** these when possible.

## C. Sorting

Arranging the clean data in some sequence and/or in different sets.

Here data is sorted on the basis of common characteristics which can either be:

- ***descriptive*** or
- ***numerical***

# Data Processing

## D. Input

In this step, the sorted raw data is converted into machine readable form and fed into the processing unit. This can be in the form of data entry through a keyboard, scanner or any other input source.

## E. Processing (Analysisng data)

Data can be analysed either **manually** or with the help of a **computer**

**I- Manual Data Analysis:** This can be done if the number of data is very small. Manual data analysis is extremely time consuming.

# Data Processing

**II- Data Analysis Using a Computer:** If you want to analyse data using computer, you should be familiar with the appropriate program. In this area, knowledge of computer and statistics plays an important role.

## F. Data output/interpretation

The output/interpretation stage is the stage at which data is finally usable to non-data scientists.

It is translated, readable, and often in the form of graphs, videos, images, plain text, etc.).

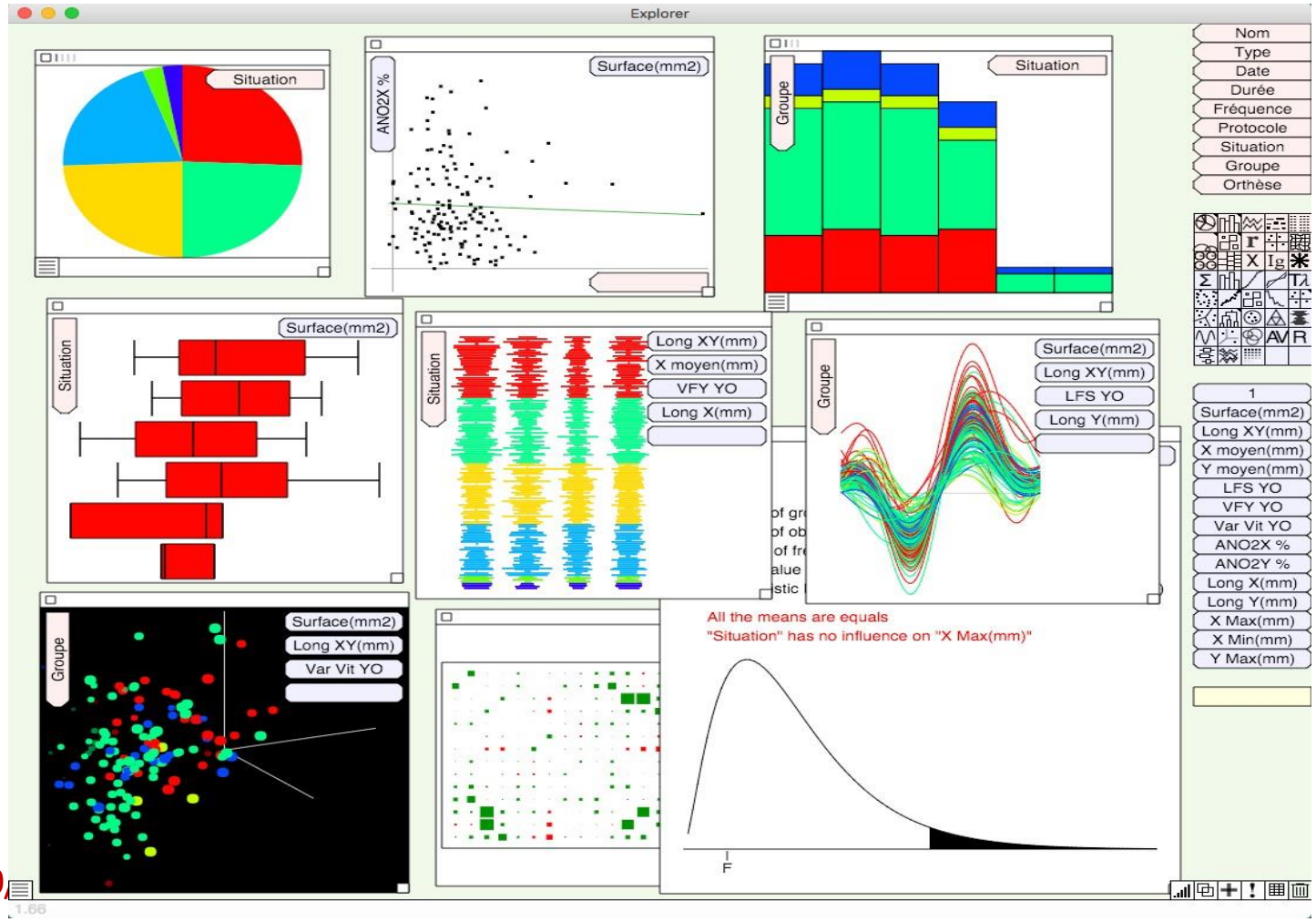
# Data Processing

## G. Storage

The last step of the data processing cycle is storage, where data and metadata are stored for further use.

This allows for quick access and retrieval of information whenever needed, and also allows it to be used as input in the next data processing cycle directly.

# Statistical analysis and its software



# Statistical analysis and its software

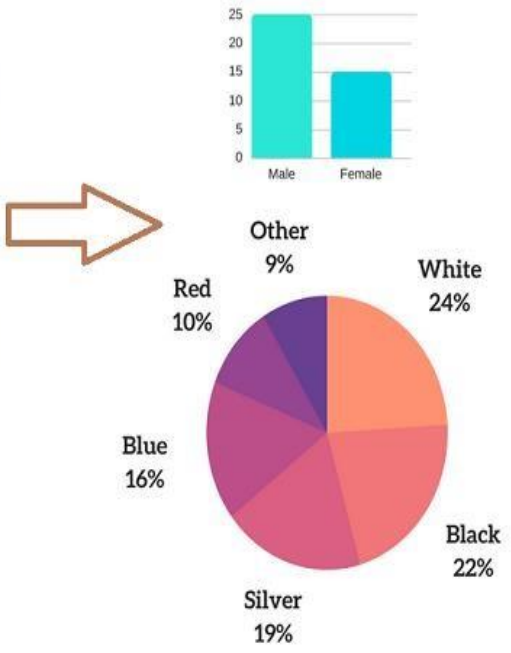
## Statistical analysis

Statistical analysis is a scientific tools that helps collect and analyze large amounts of data to identify common patterns and trends to convert them into meaningful information.

In simple words, statistical analysis is a data analysis tool that helps draw meaningful conclusions from raw and unstructured data.

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Respondent Number | Age | Gender | Favorite Car Color |
| 2 | 1 | 22 | M | White |
| 3 | 2 | 37 | F | Silver |
| 4 | 3 | 45 | F | Black |
| 5 | 4 | 62 | F | Gray |
| 6 | 5 | 28 | M | Red |
| 7 | 6 | 45 | M | Green |
| 8 | 7 | 88 | F | Brown |
| 9 | 8 | 61 | M | White |
| 10 | 9 | 95 | M | Black |
| 11 | 10 | 27 | M | White |
| 12 | 11 | 39 | F | Green |
| 13 | 12 | 43 | M | Brown |
| 14 | 13 | 55 | F | Black |
| 15 | 14 | 59 | F | White |

RAW DATA




Descriptive Statistics

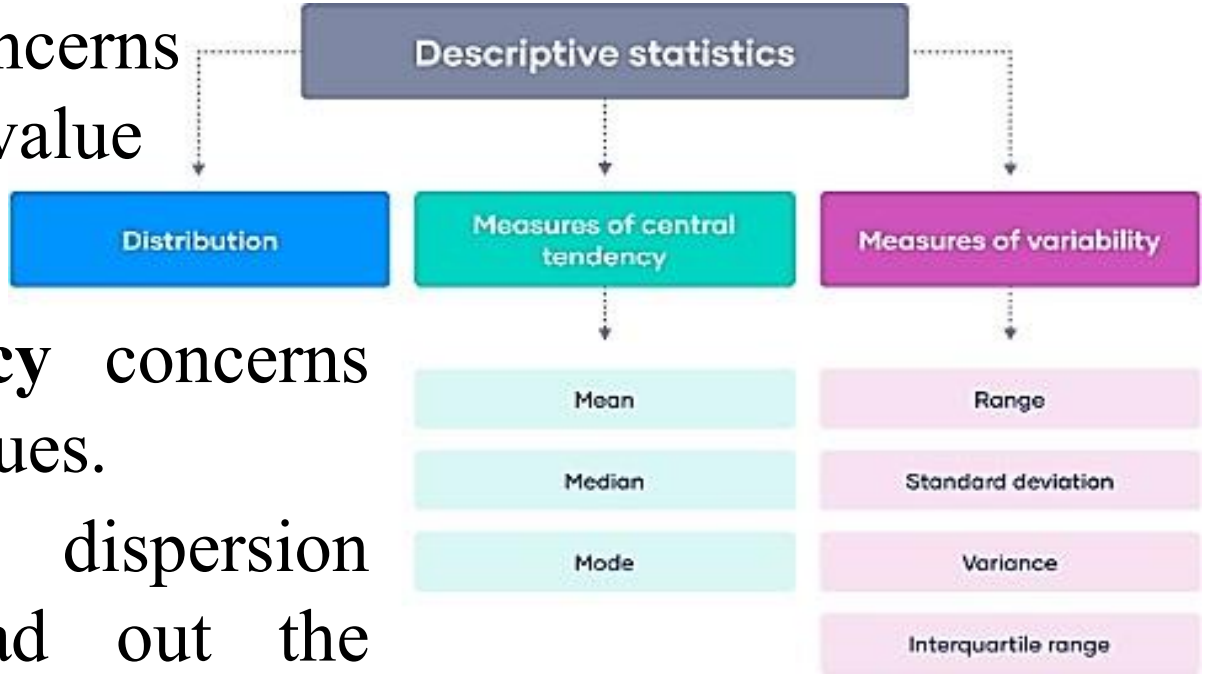
# Statistical analysis and its software

## Types of statistical analysis

### A- Descriptive statistical analysis

It used to describe the features of data and show or summarize data in the form of table, chart, and graph

The **distribution** concerns the frequency of each value




The **central tendency** concerns the averages of the values.

The **variability** or dispersion concerns how spread out the values are.

# Statistical analysis and its software

## Types of statistical analysis

### B- Inferential statistical analysis

Inferential statistics can be defined as a field of statistics that uses analytical tools for drawing conclusions about a population by examining random samples.

The goal of inferential statistics is to make generalizations about a population.

Inferential statistics can be classified into **hypothesis testing** and **regression analysis**.

# Statistical analysis and its software

## B- Inferential statistical analysis

Inferential Statistics

| Inferential Statistics | |
|---|---|
| Hypothesis Testing | Regression Analysis |
| Z test | Linear Regression |
| F test | Nominal Regression |
| T test | Logistic Regression |
| ANOVA Test | Ordinal Regression |
| Wilcoxon Signed Rank Test | |
| Mann-Whitney U Test | |

# Statistical analysis and its software

## C- Associative or relative analysis

**Associative or relative statistics** seek to identify meaningful interrelationships between or among data.

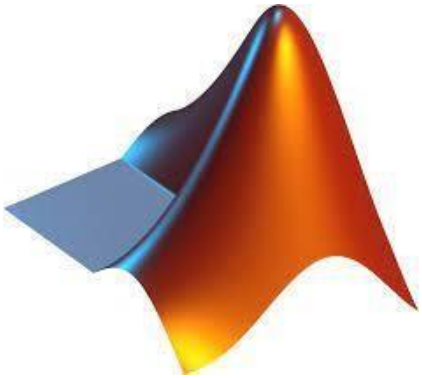
For instance, "Is there a relationship between salt intake and blood pressure among middle-age women?" is a problem definition suitable for analysis by associative statistics.

# Statistical analysis and its software

## Major statistical data analysis softwares in 2024


**Statistical Package for Social Sciences**


**Microsoft Excel**



**Matlab**


**OriginPro**


**Minitab**


**GraphPad Prism**

# Statistical analysis and its software

## Issues to consider in data analysis

There are a number of issues to consider with respect to data analysis. These include:

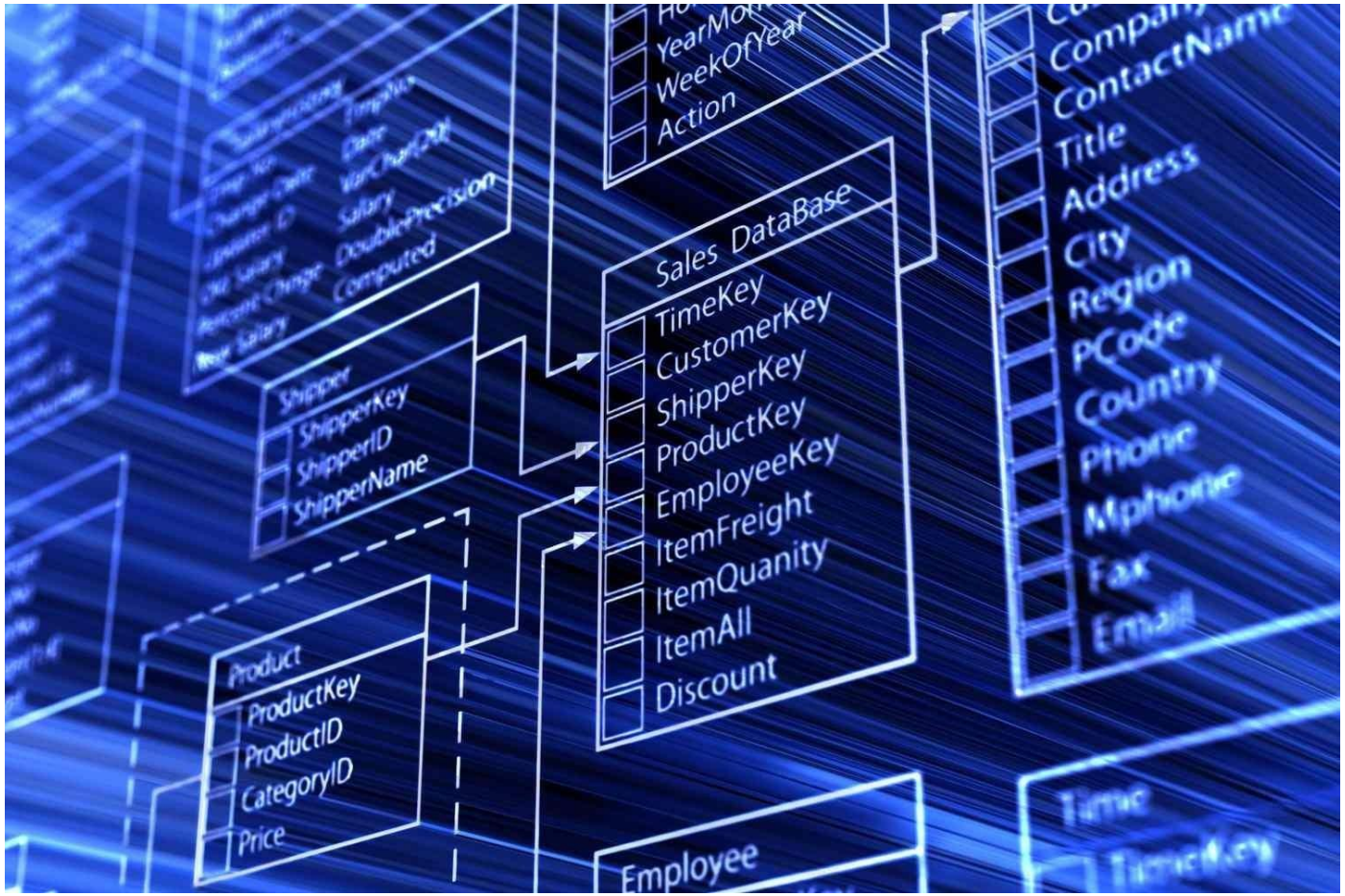
- Having the necessary skills to analyze

-Following acceptable norms for data analysis and presentation

- Choosing the appropriate statistical software

- Providing honest and accurate analysis

# Database


Sales DataBase
TimeKey
CustomerKey
ShipperKey
ProductKey
EmployeeKey
ItemFreight
ItemQuanity
ItemAll
Discount
Shipper
ShipperKey
ShipperID
ShipperName
Product
ProductKey
ProductID
CategoryID
Price
Employee
Action
Title
Address
City
Region
PCode
Country
Phone
Fax

# Database

## What is Database

A database is a collection of information that is organized so that it can easily be accessed, managed, and updated in digital form.

In one view, databases can be classified according to types of content: bibliographic, full-text, numeric, and images.

# Database

## TYPES OF DATABASE

- Relational database
- Document oriented database
- Graph database
- Hypertext database
- Operational database
- Distributed database
- Flat file database

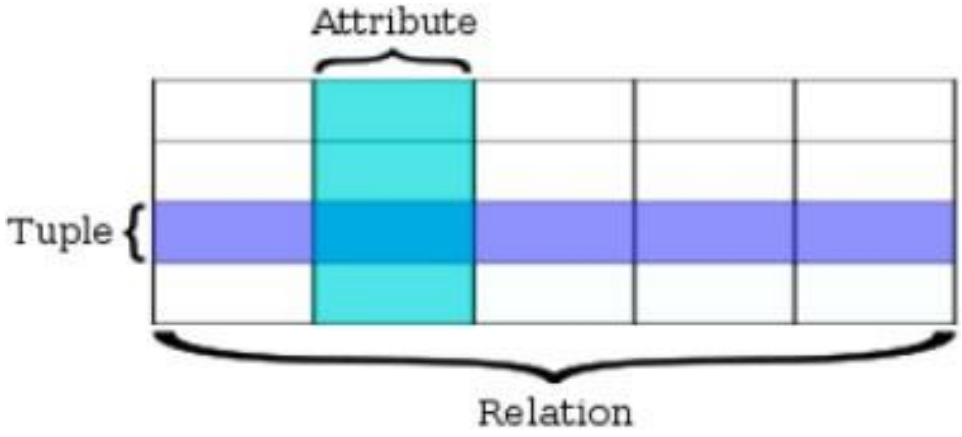
# Database

## RELATIONAL DATABASE

A relational database is a database that has a collection of tables of data items.

Data in a single table represents a relations.

Tables consist of rows and columns. The columns constitute the attributes. The rows constitute tuples.



# Database

## DOCUMENT ORIENTED DATABASE

A document-oriented database is a database that stores information in documents.

A document database is a type of **nonrelational database** that is designed to store data **as JSON-like documents**.

**A document**

```
FirstName="Bob", Address="5 Oak St.", Hobby="sailing"
```

**Another document**

```
FirstName="Jonathan", Address="15 Wanamassa Point Road", Children=("Michael,10", "Jennifer,8", "Samantha,5", "Elena,2")
```

# Database

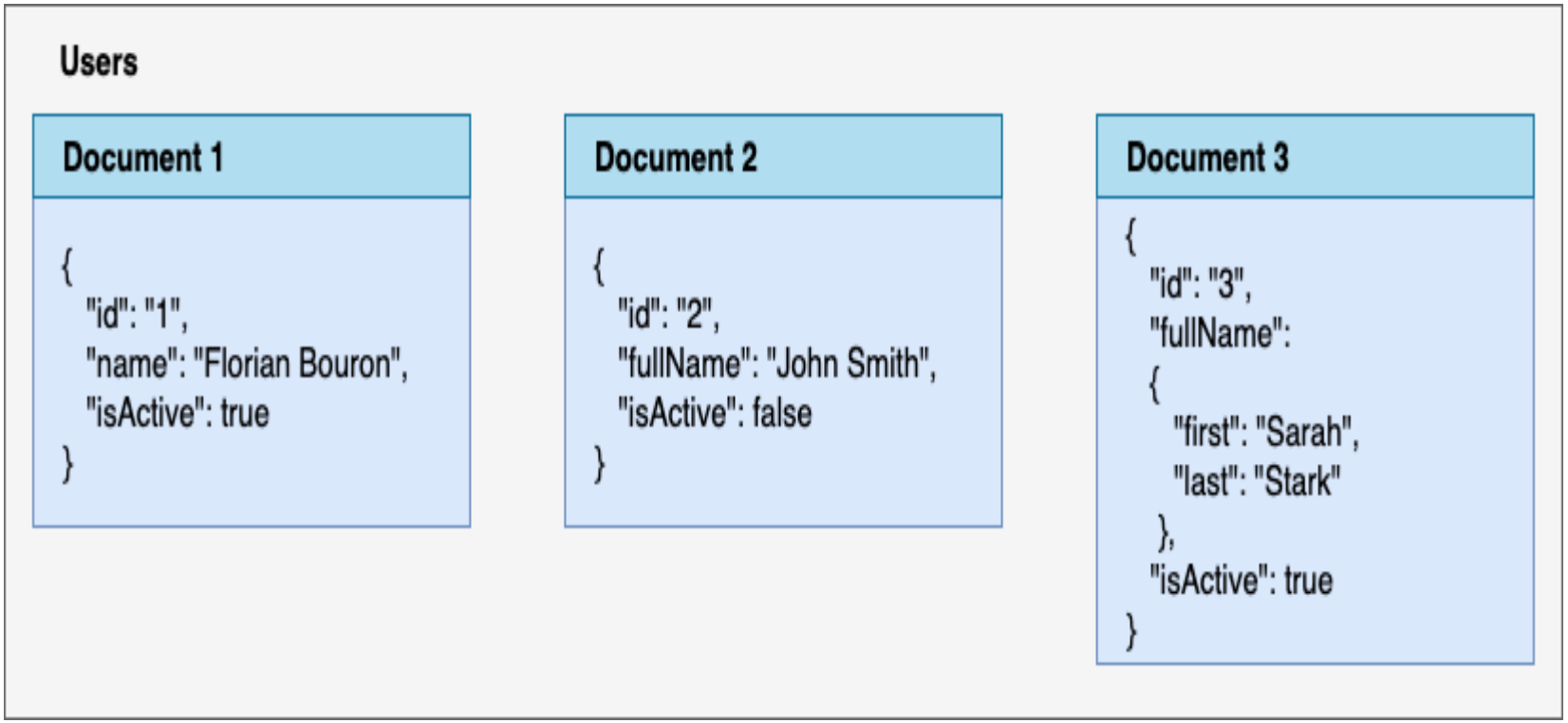
In the following example, a JSON-like document describes a book.

JSON

```
[
    {
        "year" : 2013,
        "title" : "Turn It Down, Or Else!",
        "info" : {
            "directors" : [ "Alice Smith", "Bob Jones"],
            "release_date" : "2013-01-18T00:00:00Z",
            "rating" : 6.2,
            "genres" : ["Comedy", "Drama"],
            "image_url" : "http://ia.media-imdb.com/images/N/O9ERWAU7FS797AJ7LU8HN09AMUP908RLlo5JF90EWR7LJKQ7@@._V1_SX400_.jpg",
            "plot" : "A rock band plays their music at high volumes, annoying the neighbors.",
            "actors" : ["David Matthewman", "Jonathan G. Neff"]
        }
```

# Database

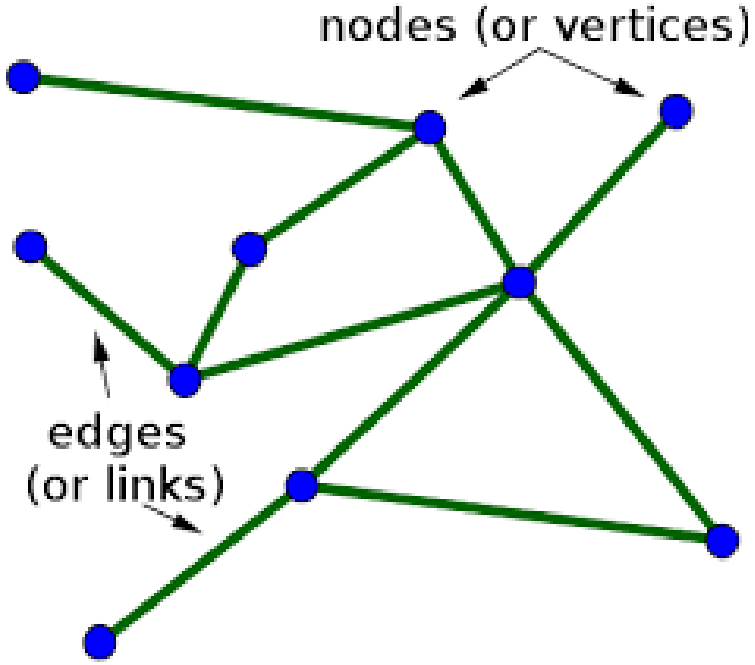
Another example



# Database

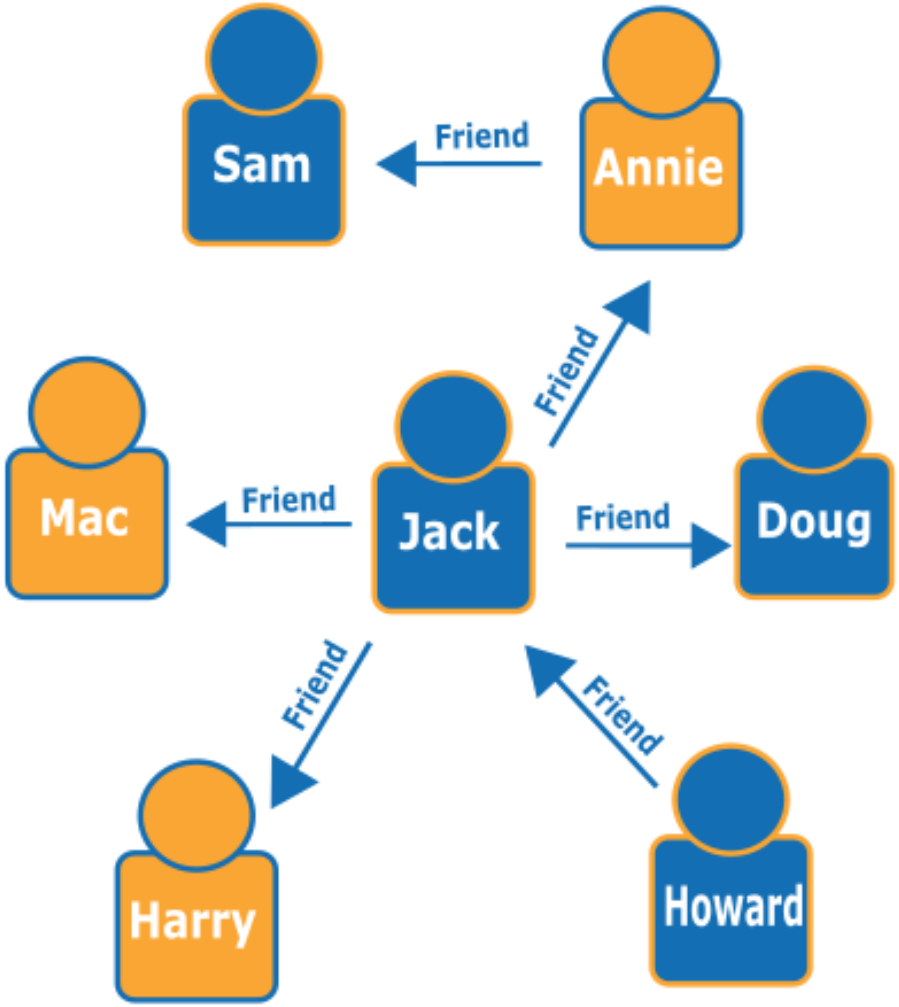
## GRAPH DATABASE

A graph database is a database that uses graph structures with **nodes, edges**, and properties to represent and store data.



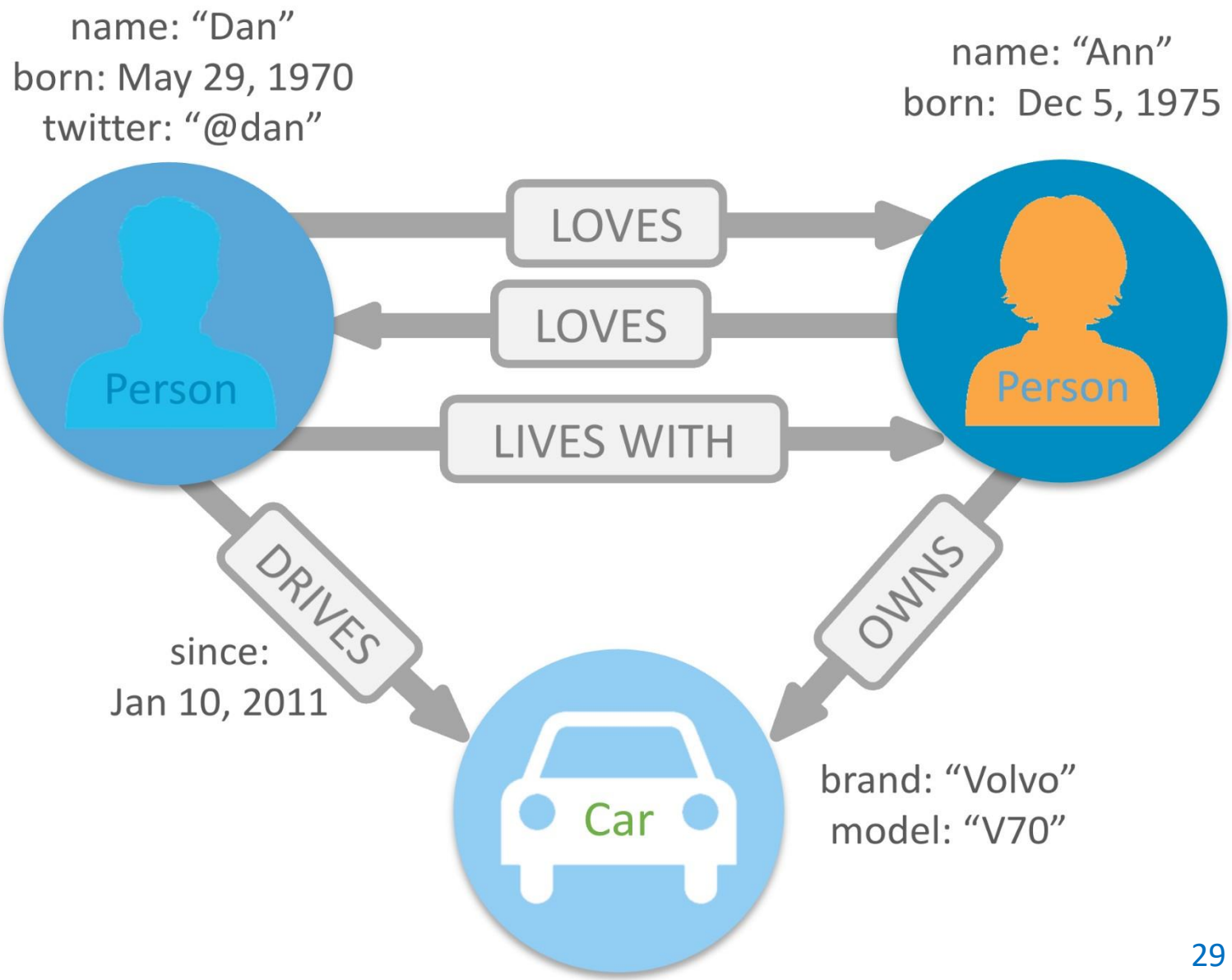
# Database

## Ex: (Facebook)



# Database

**Ex:**


name: “Dan”
born: May 29, 1970
twitter: “@dan”
name: “Ann”
born: Dec 5, 1975
LOVES
LOVES
LIVES WITH
Person
Person
DRIVES
OWNS
since:
Jan 10, 2011
Car
brand: “Volvo”
model: “V70”

# Database

## HYPERTEXT DATABASE

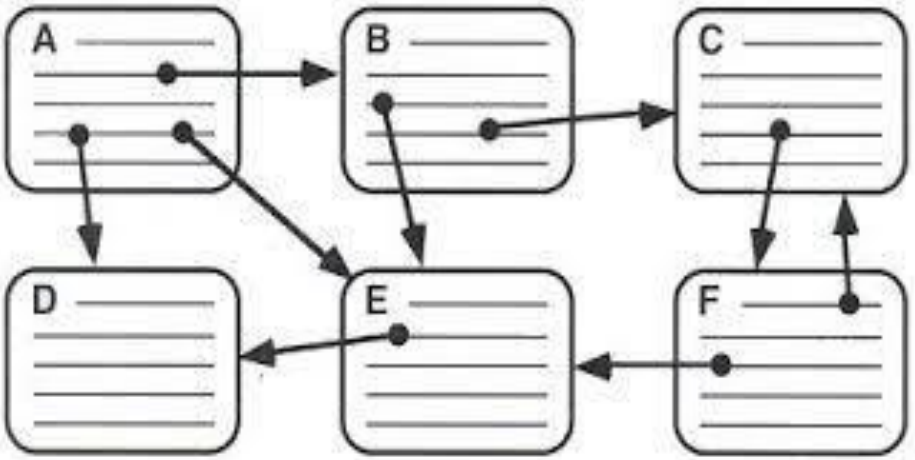
Hypertext is a special type of database system, in which **objects (text, pictures, music, programs, and so on) can be creatively linked to each other**. When you select an object, you can see all the other objects that are linked to it.

**Hypertext** is text displayed on a computer display or other electronic devices with references (hyperlinks) to other text that the reader can immediately access.

Pages often written in HTML (Hypertext Markup Language).

# Database

## Ex:




Document-oriented databases are one of the main categories of NoSQL databases, and the popularity of the term "document-oriented database" has grown[2] with the use of the term NoSQL itself. XML databases are a subclass of document-oriented databases that are optimized to work with XML documents. Graph databases are similar, but add another layer, the *relationship*, which allows them to link documents for rapid traversal.

Document-oriented databases are inherently a subclass of the key-value store, another NoSQL database concept. The difference[*contradictory*] lies in the way the data is processed; in a key-value store, the data is considered to be inherently opaque to the database, whereas a document-oriented system relies on internal structure in the *document* in order to extract metadata that the database engine uses for further optimization. Although the difference is often negligible due to tools in the systems,[a] conceptually the document-store is designed to offer a richer experience with modern programming techniques.

# Database

## OPERATIONAL DATABASE

An operational database management system is **software that is designed to allow users to easily define, modify, retrieve, and manage data in real-time.**

While conventional databases rely on batch processing, operational database systems are oriented toward real-time, transactional operations.

An operational database contains data about the things that go on inside an organization.

Example: data on customer complaints, employee information, etc..

# Database

## DISTRIBUTED DATABASE

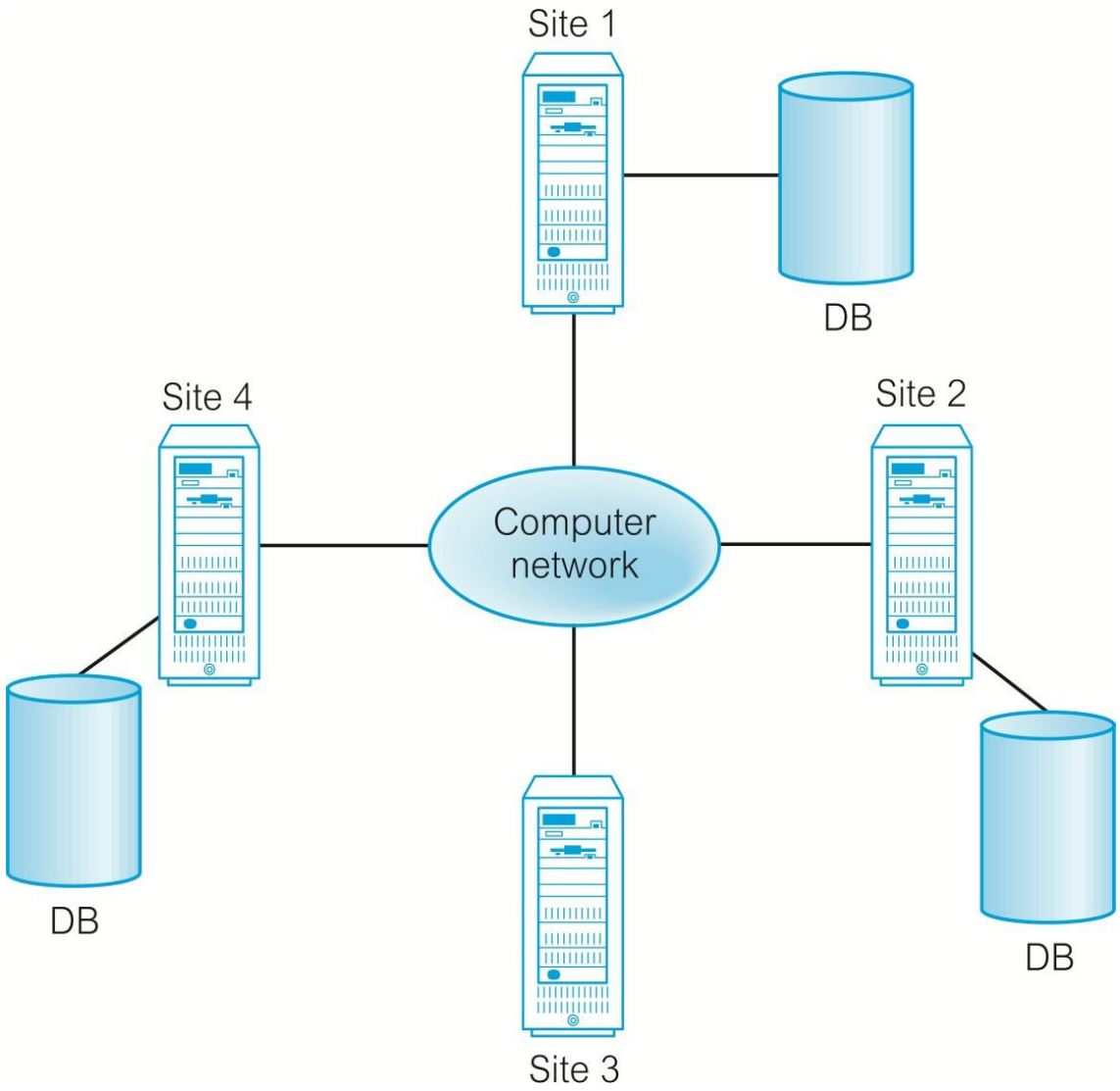
It is a database in which portions of the database are stored on **multiple computers** within a network or different networks.

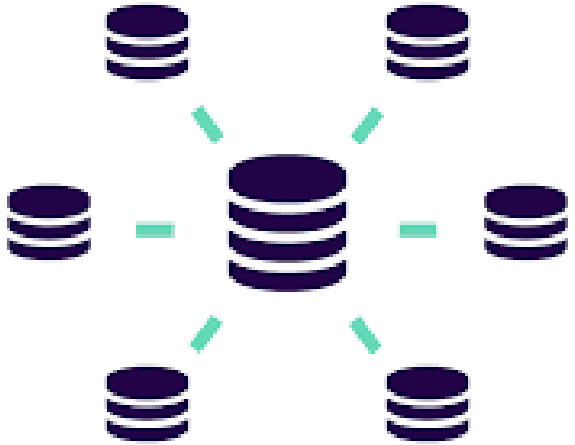
A distributed database is basically a database that **is not limited to one system, it is spread over different sites, i.e, on multiple computers or over a network of computers.**

Updates and deletes performed on the data at one location will be automatically reflected in the data stored elsewhere.

# Database

**Ex:**



# Database

## FLAT FILE DATABASE

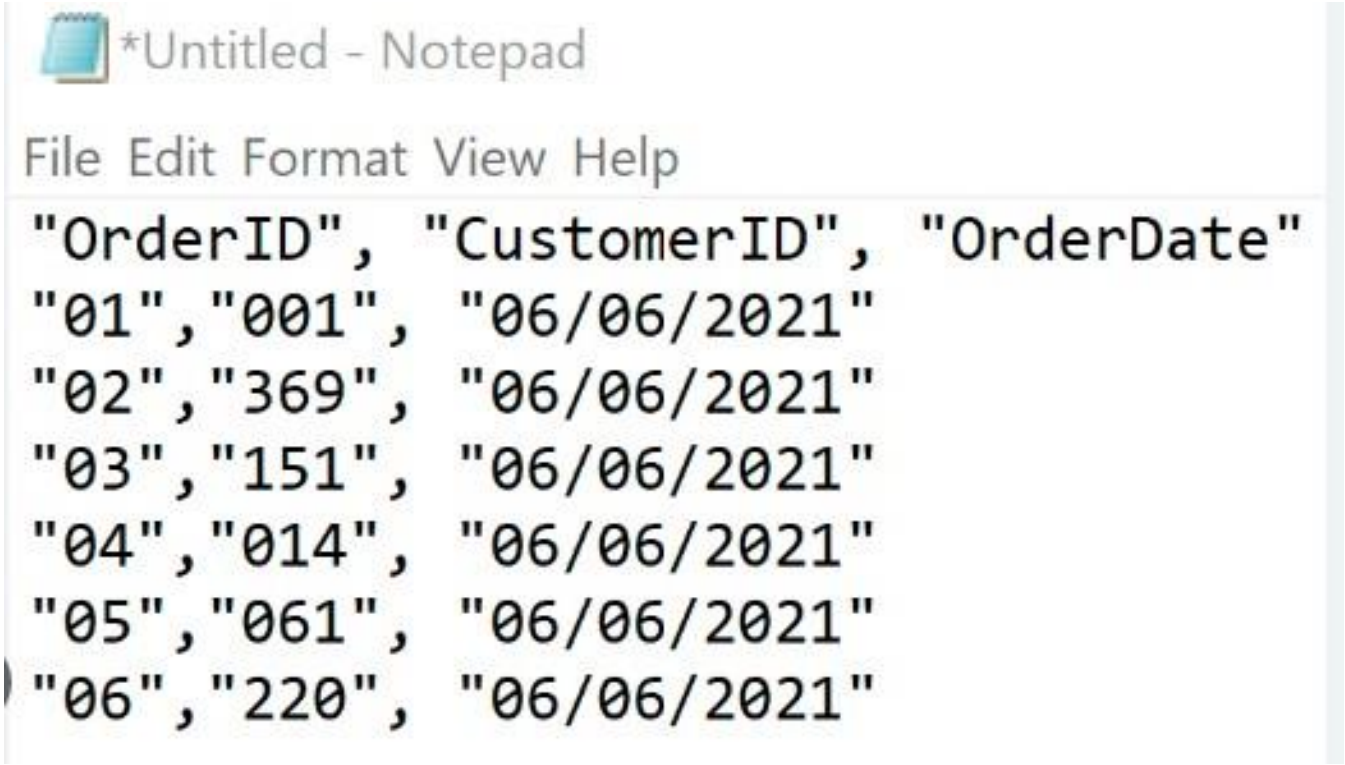
A flat file database is a type of database that stores data in a **single table**. This is unlike a relational database, which makes use of multiple tables and relations.

Flat file databases are generally in **plain-text form**, where each line holds only one record, , with fields separated by commas or tabs.

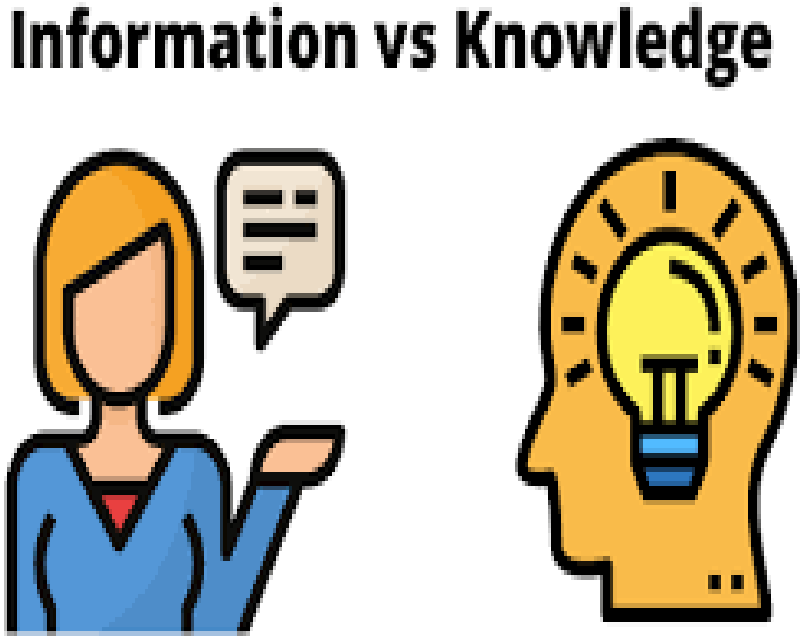
Flat files are ideal for small amount of data.

# Database

**Ex:**


*Untitled - Notepad
File Edit Format View Help
"OrderID", "CustomerID", "OrderDate"
"01","001", "06/06/2021"
"02","369", "06/06/2021"
"03","151", "06/06/2021"
"04","014", "06/06/2021"
"05","061", "06/06/2021"
"06","220", "06/06/2021"

# Knowledge



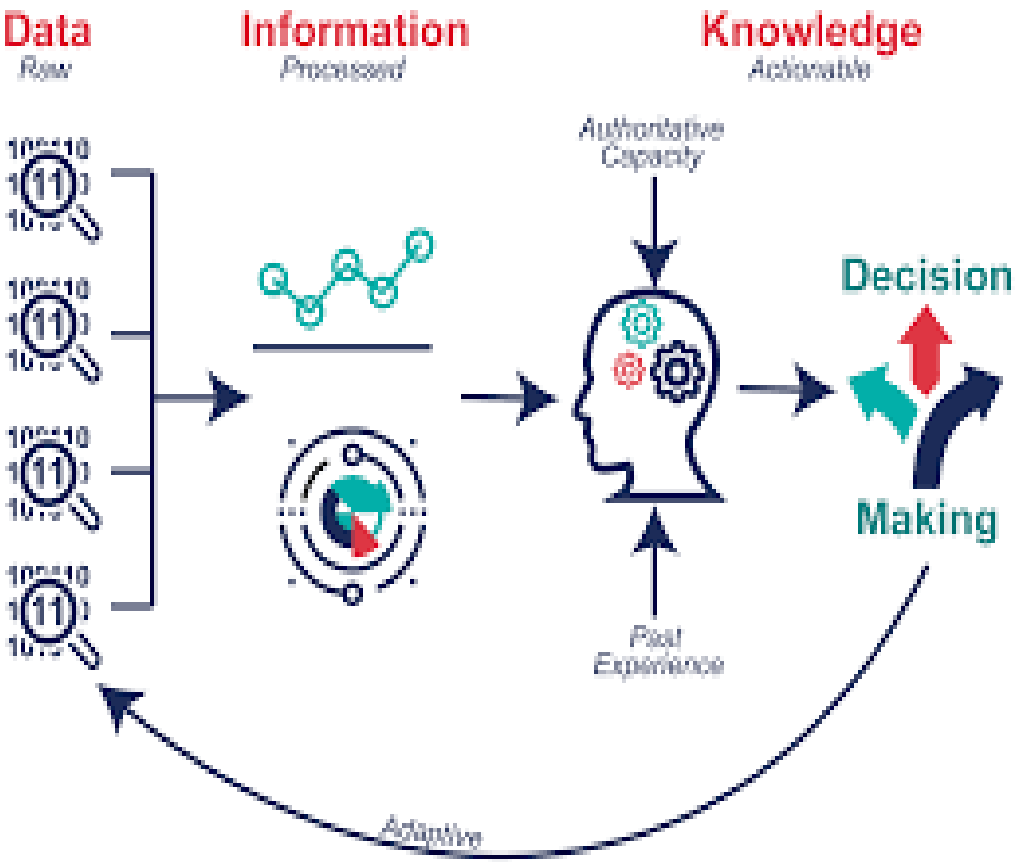
# Data & Information and Knowledge

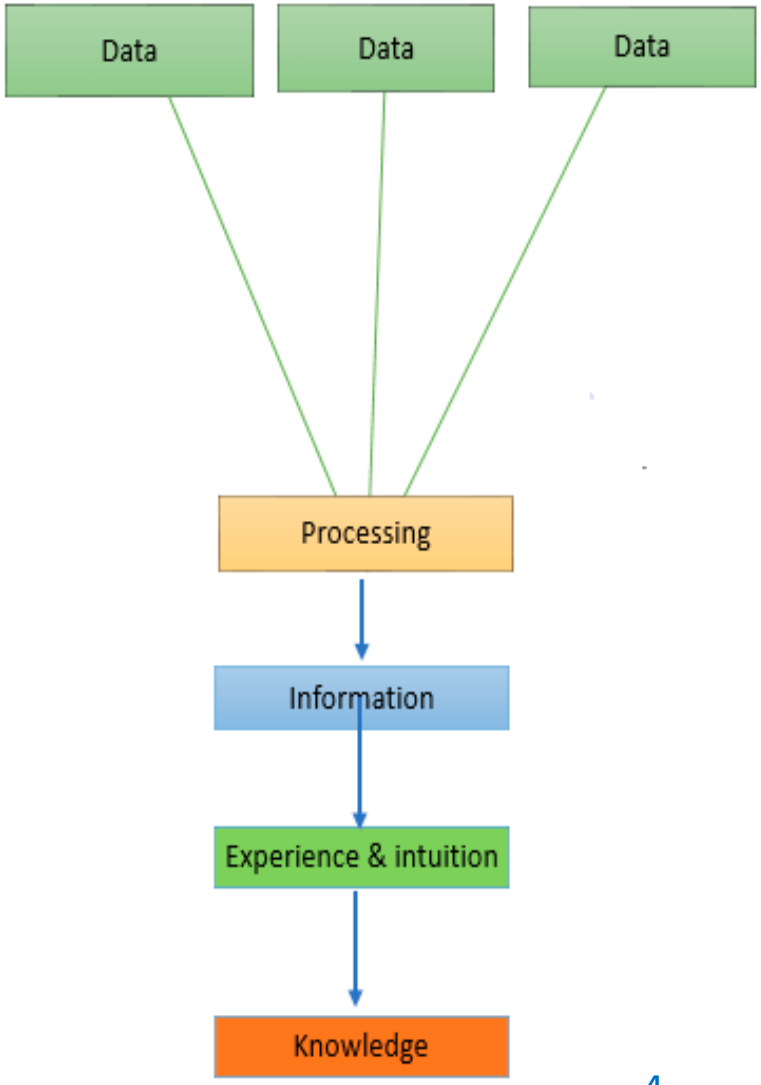
-**Knowledge** is a combination of information, experience, and insight that helps the individual or the organization for linking to doing and implies know-how and understanding.

-**Knowledge** is possessed by each individual and is an outcome of his or her experience. It also covers the norms to evaluates new inputs from his surroundings

- **Information** is a text that answers the questions of **who**, **when**, or **where**, while **knowledge** is a text that answers the questions of **why** and **how**.

# Difference Between Information and Knowledge



# Data & Information and Knowledge

## Knowledge

- Using the temperature data and information provided earlier, knowledge might involve understanding:

- Weather patterns,

- How temperature changes throughout the day,

-The potential impact of such changes on various aspects like human comfort or plant growth.

This understanding can be applied to predict future temperature trends or decide when to engage in certain outdoor activities.

| BASIS FOR COMPARISON | INFORMATION | KNOWLEDGE |
|---|---|---|
| **Combination of** | Data processing and context | Information, experience and intuition |
| **Processing** | Improves representation | Increases awareness |
| **Transfer** | Easily transferable | Requires learning |
| **Prediction** | Information alone is not sufficient to make predictions | Prediction is possible if one possess required knowledge. |
| **One in other** | Not all information is knowledge. | All knowledge is information. |