**🟩 What is Data?**

Data refers to any collection of facts, figures, or information that can be recorded, measured, or analyzed.

**➤ Types of Data by Structure:**

**1️⃣ Structured Data**

* **Definition**: Data that is organized in a predefined format, typically rows and columns (like in spreadsheets or databases).
* **Examples**:
  + Excel tables with student records: Name | Age | Marks
  + SQL databases: Employee\_ID, Salary, Department
  + Sensor logs with timestamps
* **Use Cases**:
  + Ideal for business analytics, machine learning models, dashboards
* **Why It’s Needed**: Easy to store, query, and visualize using tools like Python, SQL, Power BI.

**2️⃣ Unstructured Data**

* **Definition**: Data that doesn't have a specific format or structure. Harder to organize and analyze directly.
* **Examples**:
  + Text documents (emails, social media posts)
  + Images (X-rays, product photos)
  + Audio/video recordings (interviews, podcasts)
* **Use Cases**:
  + Used in NLP (text), computer vision (images), speech recognition
* **Why It’s Needed**: Real-world data is mostly unstructured; valuable insights are hidden in it.

**🟨 Data Categories**

Let’s now classify data based on its nature:

**➤ Data → Numerical vs Non-Numerical**

**🔢 1. Numerical Data (Quantitative)**

Values that can be measured or counted.

**a) Discrete Data**

* Countable, whole numbers.
* Examples:
  + Number of students in a class
  + Cars in a parking lot
* **Visuals**: Bar charts, pie charts

**b) Continuous Data**

* Measurable, can take infinite values within a range.
* Examples:
  + Height of students (in cm)
  + Temperature, weight
* **Visuals**: Line graphs, histograms

**🔠 2. Non-Numerical Data (Qualitative)**

**a) Categorical Data**

Represents categories or labels.

**i) Nominal Data**

* No logical order or ranking.
* Examples:
  + Gender (Male, Female, Other)
  + City names
* **Visuals**: Bar chart, pie chart

**ii) Ordinal Data**

* Categories with a logical order.
* Examples:
  + Education Level (High School, Bachelor, Master)
  + Customer Satisfaction (Poor, Average, Good, Excellent)
* **Visuals**: Ordered bar chart

**b) Non-Categorical Free Text (Semi-structured)**

* User comments, descriptions
* Not easily visualized without NLP
* Examples:
  + Feedback: "The product was amazing!"

**📝 Summary Table**

| **Type** | **Subtype** | **Ordered?** | **Example** | **Suitable Charts** |
| --- | --- | --- | --- | --- |
| Numerical | Discrete | No | Number of children | Bar chart, pie chart |
| Numerical | Continuous | Yes | Weight, height | Line graph, histogram |
| Non-Numerical | Nominal | No | Gender, colors | Bar chart, pie chart |
| Non-Numerical | Ordinal | Yes | Ratings (Low, Medium, High) | Ordered bar chart |

**🟪 Dimensions of Data (Tensor Perspective)**

Data can be represented as **tensors**, where dimensions indicate the structure and complexity of the data. Here’s a breakdown from **1D to 5D**, with real-world examples:

**🔹 1D – One-Dimensional Data**

* **Structure**: A single list or array
* **Shape Example**: (n,)
* **Example**:
  + A sentence represented as a sequence of word embeddings (text → [x1, x2, x3, ..., xn])
  + Audio waveform (raw samples over time)
* **Use Case**: NLP, signal processing
* **Visualization**: Line graph

**🔹 2D – Two-Dimensional Data**

* **Structure**: Rows and columns (matrix)
* **Shape Example**: (rows, columns)
* **Example**:
  + Tabular data (CSV files: rows = samples, columns = features)
  + Grayscale image → pixel grid (Height × Width)
* **Use Case**: Statistical data, BW image processing
* **Visualization**: Heatmaps, tables

**🔹 3D – Three-Dimensional Data**

* **Structure**: Stack of 2D data (like depth)
* **Shape Example**: (channels, height, width) or (height, width, channels)
* **Example**:
  + RGB image: 3 color channels (Red, Green, Blue)
  + Time-series data across multiple sensors
* **Use Case**: Image processing, convolutional neural networks
* **Visualization**: 3D plots (limited), image viewers

**🔹 4D – Four-Dimensional Data**

* **Structure**: Multiple 3D data items (like batches)
* **Shape Example**: (batch, height, width, channels)
* **Example**:
  + Video: sequence of RGB frames over time  
    (frames × height × width × channels)
  + Image dataset used in deep learning
* **Use Case**: Video analysis, training image models
* **Visualization**: Frame-by-frame viewer

**🔹 5D – Five-Dimensional Data**

* **Structure**: Sequence of videos or videos with additional context like audio or batch
* **Shape Example**: (batch, time, height, width, channels)
* **Example**:
  + Video with audio (each video is a sequence of RGB frames, aligned with audio spectrograms)
  + 3D medical imaging over time (MRI scans over frames for multiple patients)
* **Use Case**: Multimodal learning, spatiotemporal AI
* **Visualization**: Complex and usually broken down

**✅ Is Your Mapping Correct?**

Yes, your understanding is largely **correct** with a few clarifications:

| **Your View** | **Adjusted Version (More Precise)** |
| --- | --- |
| 1D → Text | ✅ Yes (word or character embeddings = 1D vectors) |
| 2D → Rows, Columns / BW Image | ✅ Yes (Matrix form, grayscale image) |
| 3D → RGB Image | ✅ Yes (Height × Width × 3 channels) |
| 4D → Video | ✅ Yes (Sequence of RGB images over time) |
| 5D → Video + Audio | ✅ Yes, though audio may be handled as a separate modality; combined as 5D when synchronized |