UNIT II - UNDERSTANDING DATA AND PATTERN RECOGNITION

- Performing analytics on numeric data using any spreadsheet software and visualizing the data using charts, histograms, scatter plots, graphs
- Logical thinking – reasoning, Pattern recognition in data, data sequences, puzzles, nonograms.
- Data Encryption – ciphering sentences and Compression.

WHAT IS DATA ?

- Data is a collection of raw, **unprocessed facts, figures, symbols, or observations** that have not yet been organized or given meaning.

- Data is essential because it transforms raw facts into actionable insights that drive smart decisions, problem-solving, and continuous growth



Types of Data

Data can be broadly classified into two main types

Qualitative Data

Data that describes qualities or characteristics, not numbers.

Nominal Data

Categorical data with no natural order or ranking.

Examples:

- Blood group (A, B, AB, O)
- Programming language (Python, Java, C++)

Ordinal Data

Categorical data with a meaningful order or ranking.

Examples:

- Satisfaction level (Low, Medium, High)
- Class rank (1st, 2nd, 3rd)



Quantitative Data

Data that represents quantities and is expressed in numbers.

Discrete Data

Numeric data that can take only specific, countable values (usually whole numbers).

Examples:

- Number of students in a class
- Number of errors in a program

Continuous Data

Numeric data that can take any value within a range (including decimals).

Examples:

- Height (e.g., 165.5 cm)
- Temperature (e.g., 28.6 °C)

EXAMPLE DATASET - STUDENT PERFORMANCE

Student	Study Hours	Attendance %	Assignment	Midterm	Final
A	5	92	82	76	84
B	2	78	65	61	63
C	7	96	91	88	93
D	4	85	74	70	72
E	9	98	95	92	96
F	3	81	68	64	66
G	6	90	86	80	84
H	8	94	89	87	91

PERFORMING ANALYTICS ON NUMERICAL DATA

Measure	What it means	Formula (Excel)	Example using the data
Mean (Average)	The sum of all values divided by the number of values.	`=AVERAGE(range)`	`=AVERAGE(F2:F9)`
Median	The middle value when the data is arranged in order.	`=MEDIAN(range)`	`=MEDIAN(F2:F9)`
Mode	The value that occurs most often in the data.	`=MODE.SNGL(range)`	`=MODE.SNGL(F2:F9)`
Maximum	The highest value in the data.	`=MAX(range)`	`=MAX(F2:F9)`
Minimum	The lowest value in the data.	`=MIN(range)`	`=MIN(F2:F9)`

PERFORMING ANALYTICS ON NUMERICAL DATA

Measure	What it means	Formula (Excel)	Example using the data
Range	The difference between the highest and lowest values.	`=MAX(range)-MIN(range)`	`=MAX(F2:F9)-MIN(F2:F9)`
Count	The number of numerical values in the data.	`=COUNT(range)`	`=COUNT(F2:F9)`
Standard Deviation	Shows how spread out the values are from the mean.	`=STDEV.S(range)`	`=STDEV.S(F2:F9)`
Variance	Measures how much the values vary from the mean.	`=VAR.S(range)`	`=VAR.S(F2:F9)`

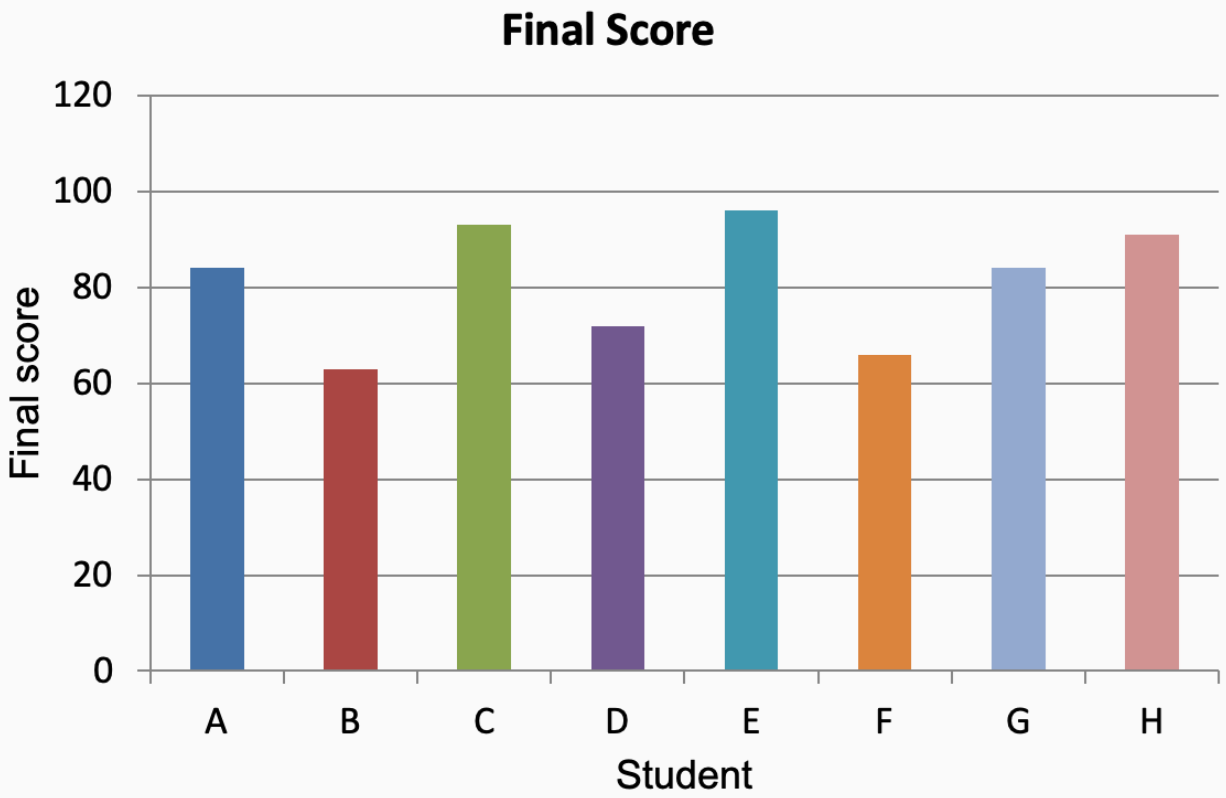
SCREENSHOT

Student	Study Hours	Attendance %	Assignment	Midterm	Final
A	5	92	82	76	84
B	2	78	65	61	63
C	7	96	91	88	93
D	4	85	74	70	72
E	9	98	95	92	96
F	3	81	68	64	66
G	6	90	86	80	84
H	8	94	89	87	91
				MEAN	81.125
				MEDIAN	84
				MODE	84
				MAXIMUM	96
				MINIMUM	63
				RANGE	33
				COUNT	8
				STD	12.631452
				VARIANCE	159.55357

VISUALISING THE DATA

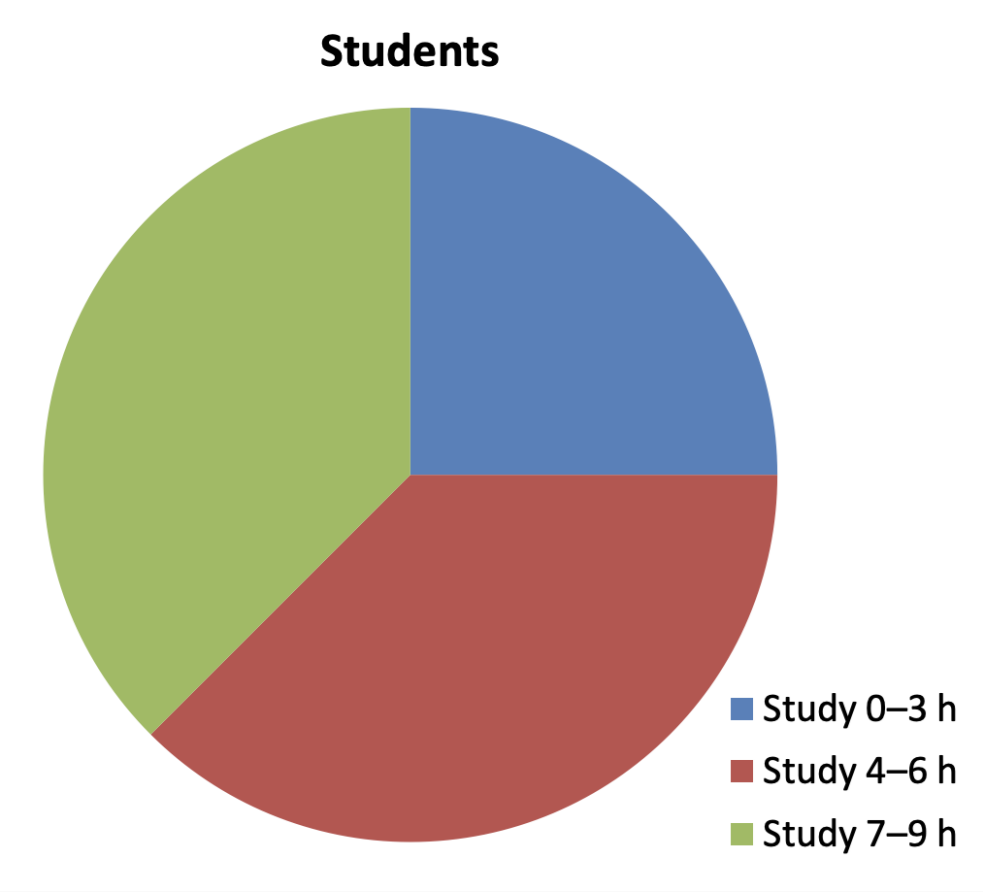
- Data visualization is the **graphical representation of information and data.**
- By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to **see and understand trends**, outliers, and patterns in data.
- Additionally, it provides an excellent way for employees or business owners to **present data to non-technical audiences** without confusion.

BAR CHART - Compare Categories



- Best for **comparing discrete categories**.
- Easy to **identify highest and lowest values**.
- Keep category labels short and consistent.

PIE CHART - Show part to whole composition



- Use when categories represent parts of one meaningful whole.
- **Keep the number of slices small.**
- Avoid using pie charts to show trends over time.

LINE CHART - Show trends over time

Month	Average Score
January	68
February	71
March	74
April	78
May	82
June	86

Enter the data in Excel

A1 = Month

B1 = Average Score

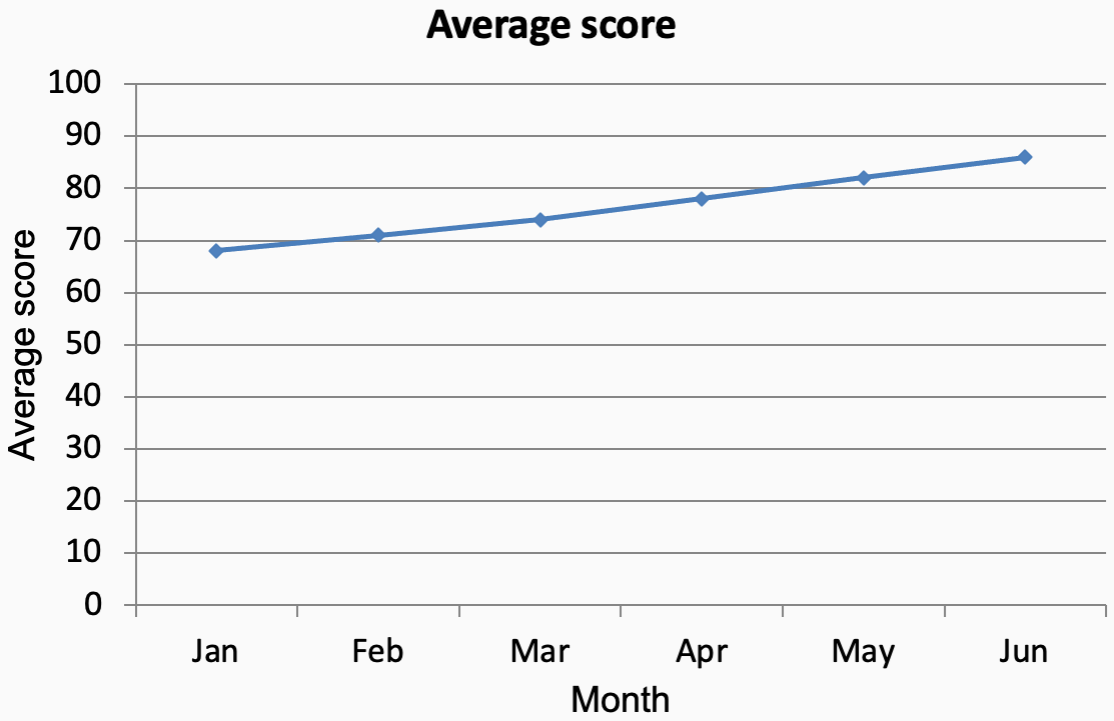
A2:A7 = January to June

B2:B7 = 68, 71, 74, 78, 82, 86

Create the line chart

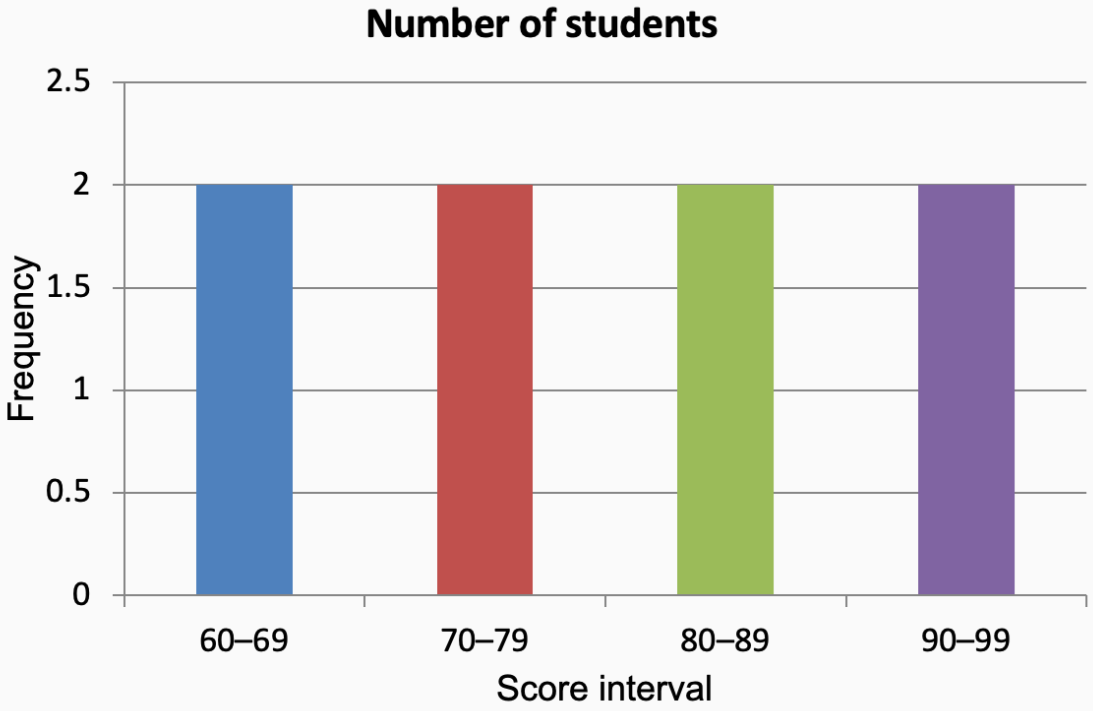
1. Select A1:B7
2. Click Insert
3. Select Line Chart
4. Title it “Monthly Average Score”

LINE CHART - Show trends over time



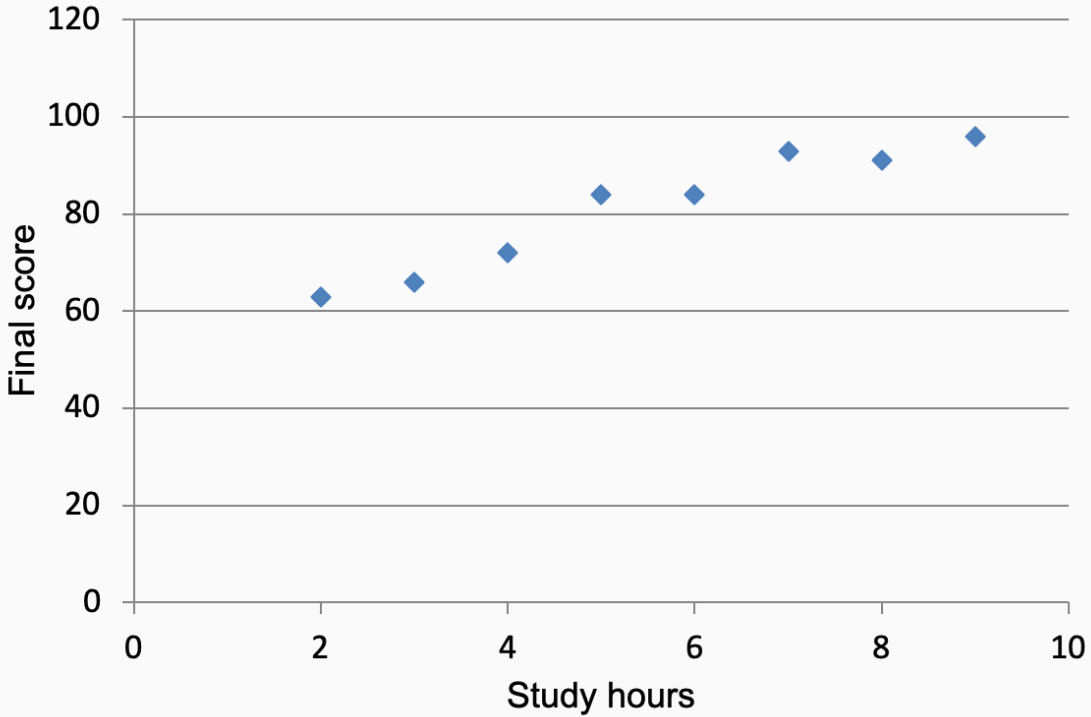
- **Use for ordered data where sequence matters.**
- Excellent for **time-series data.**
- Look for upward/ downward trends, peaks and sudden changes.

HISTOGRAM - Understand a numeric distribution



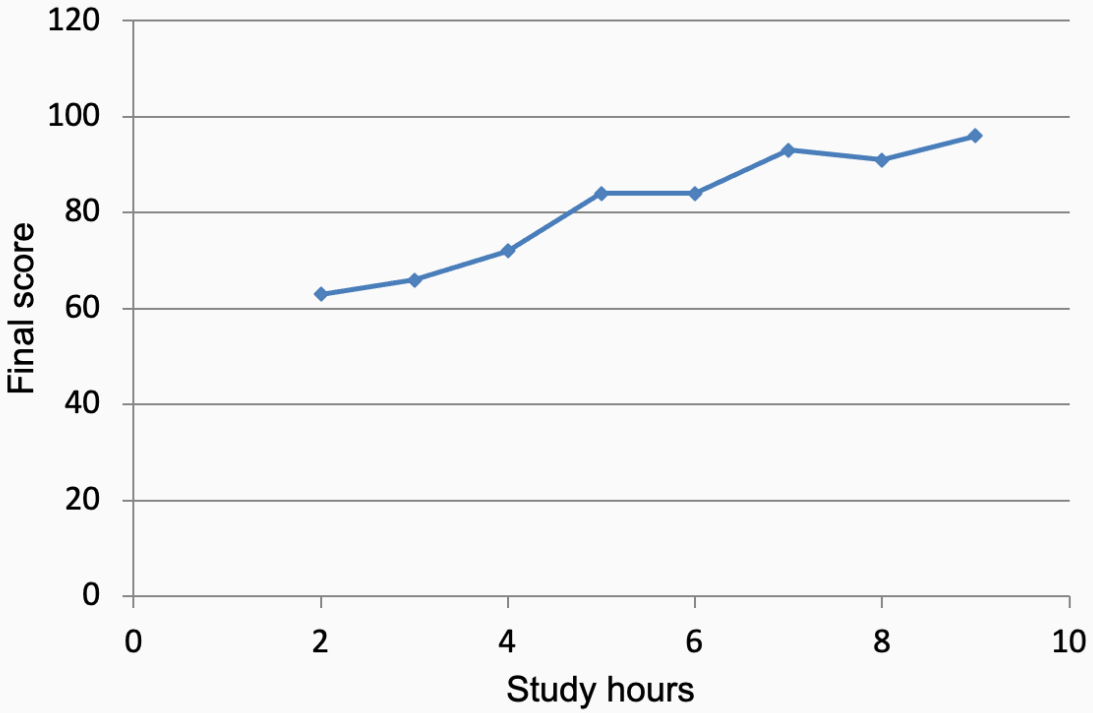
- A histogram **groups continuous numeric values into bins.**
- It reveals shape: concentration, spread and possible outliers.
- Bin width can change the story—choose it carefully.

SCATTER PLOT - Explore relationships



- **Each point represents one observation.**
- Use it to investigate relationships between two numeric variables.
- A visible upward pattern suggests positive association.

GRAPH - Plot numeric data on X-Y Axes



- X-axis: independent/input variable.
- Y-axis: dependent/output variable.
- Each point represents an observation.
- A graph can reveal trends, patterns, gaps and unusual observations.

Which visualization should you choose ?

Analytical question	Recommended chart
How do categories compare?	Bar chart
How is one total divided?	Pie chart
How does a value change over time?	Line chart
How are numeric values distributed?	Histogram
Are two numeric variables related?	Scatter plot

Good visualization practices

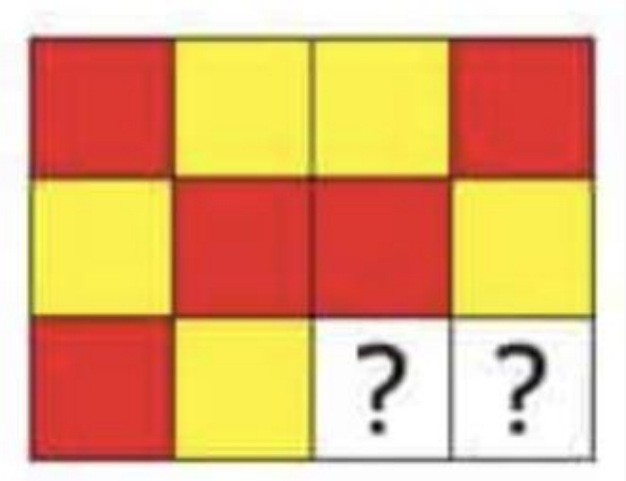
- Give every chart a clear, informative title.
- Label axes and include units where relevant.
- Use consistent scales and avoid misleading axis ranges.
- Remove unnecessary decoration and 3D effects.
- Use color intentionally: highlight the key comparison, not everything.
- Sort categories when ranking is important.
- Add data labels only when they improve readability.

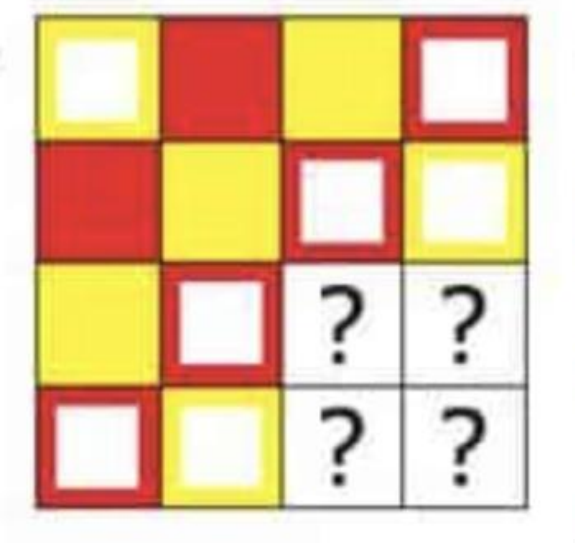
For more charts, graphs - The Data Visualisation Catalogue

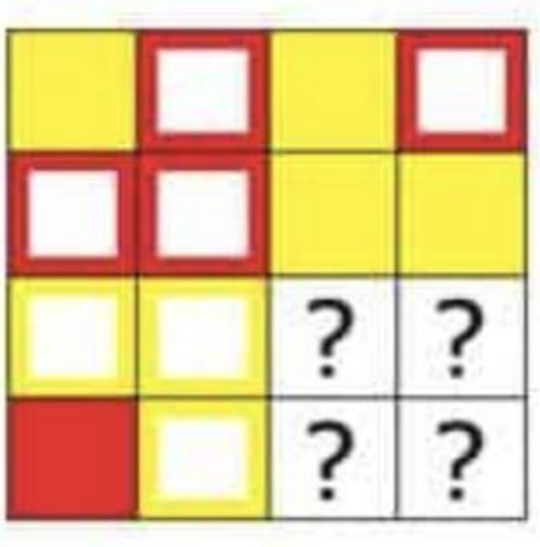
Logical Thinking and Reasoning

- **Logical thinking is the process of reasoning consistently to reach a conclusion.**
- Put simply, logical thinking means following a chain of reason, step by step, to reach a valid conclusion.
- Problems or situations involving logical thinking call for structure, relationships between facts, and chains of reasoning that “make sense.”

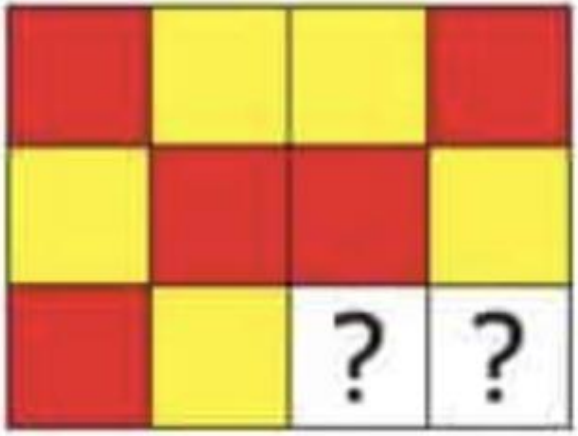
Replace the question marks with either a solid yellow, solid red, yellow border or red border

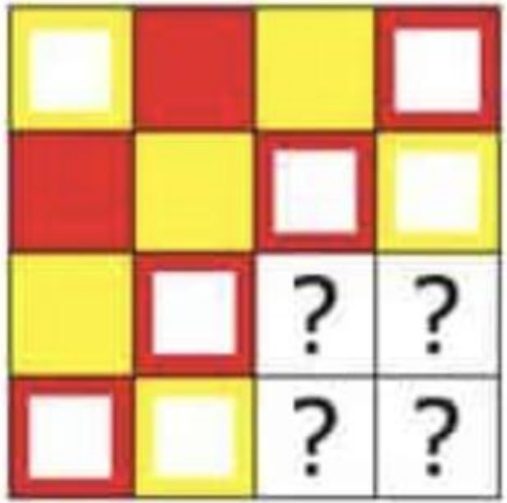


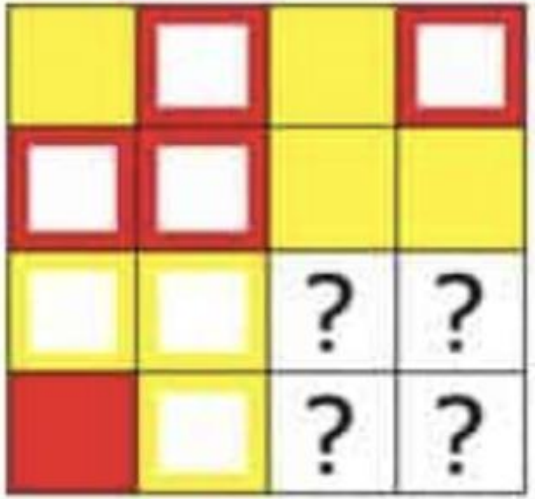


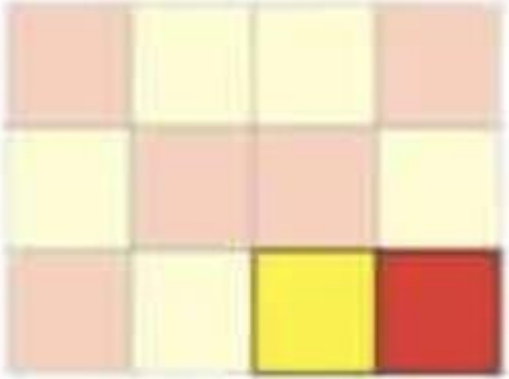


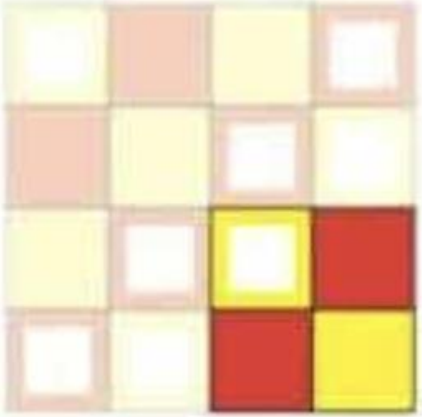
Ans in next page

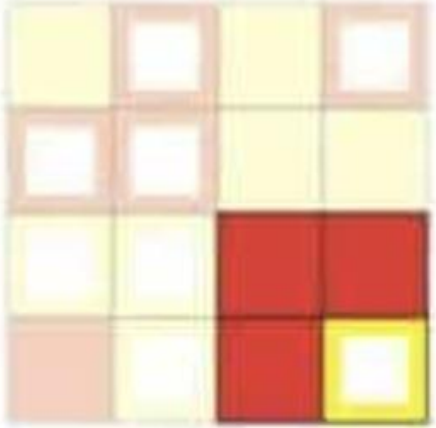












Types of Reasoning

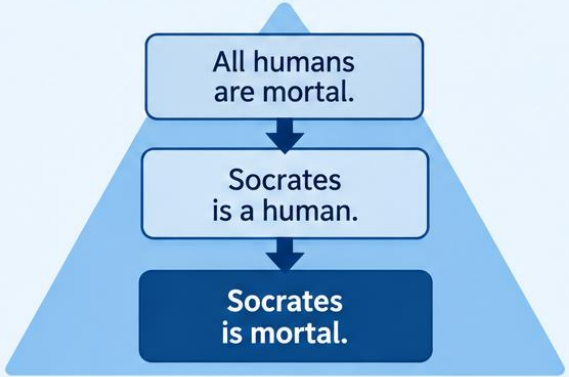
Deductive Reasoning

General Rule → Specific Conclusion

Starts with a general statement and uses logical rules to reach a specific conclusion.

Example:

All humans are mortal.
Socrates is a human.
Therefore, Socrates is mortal.



If the premises are true, the conclusion must be true.

Inductive Reasoning

Specific Observations → General Conclusion

Starts with specific observations and looks for a pattern to reach a general conclusion.

Example:

The sun has risen every day this week.
The sun has risen every day last week.
The sun has risen every day for many years.
Therefore, we conclude that the sun will rise tomorrow.



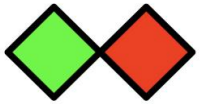
The conclusion is likely true, but not guaranteed.

Pattern Recognition

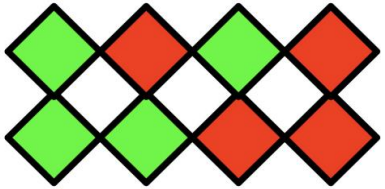
- The process of **looking for similarities, trends, and repeated structures within or across problems.**
- Once a pattern is spotted, you can predict what comes next, or reuse a solution you already know.
- Used in weather forecasting, stock market trends, fraud detection in banking

1. Find the pattern

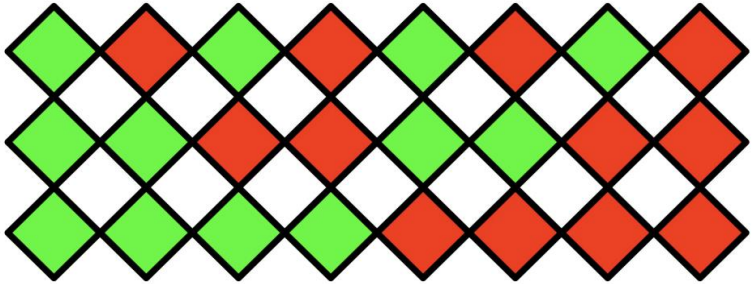
1



2



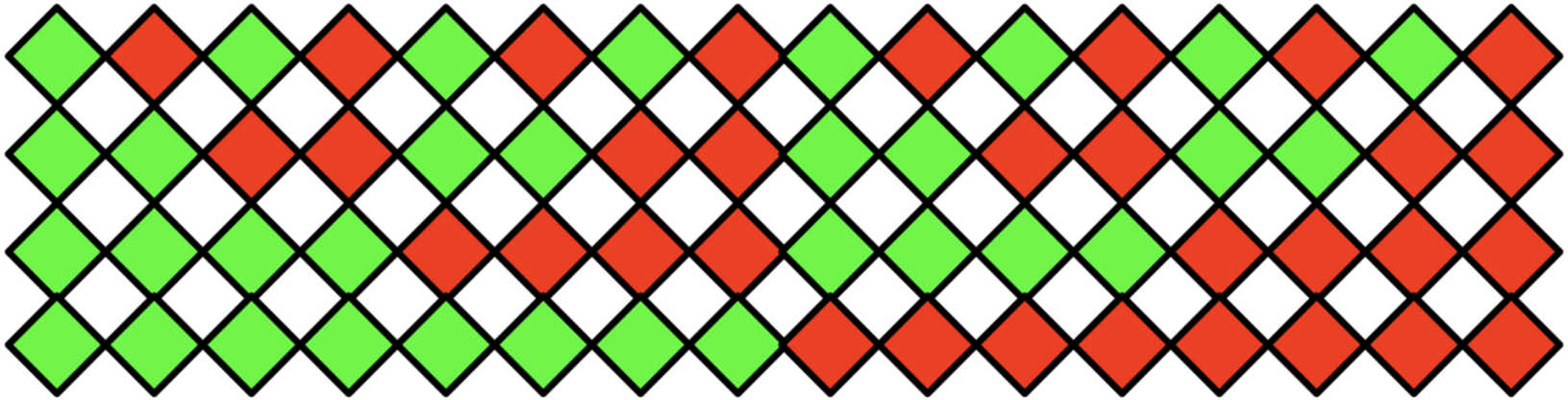
3



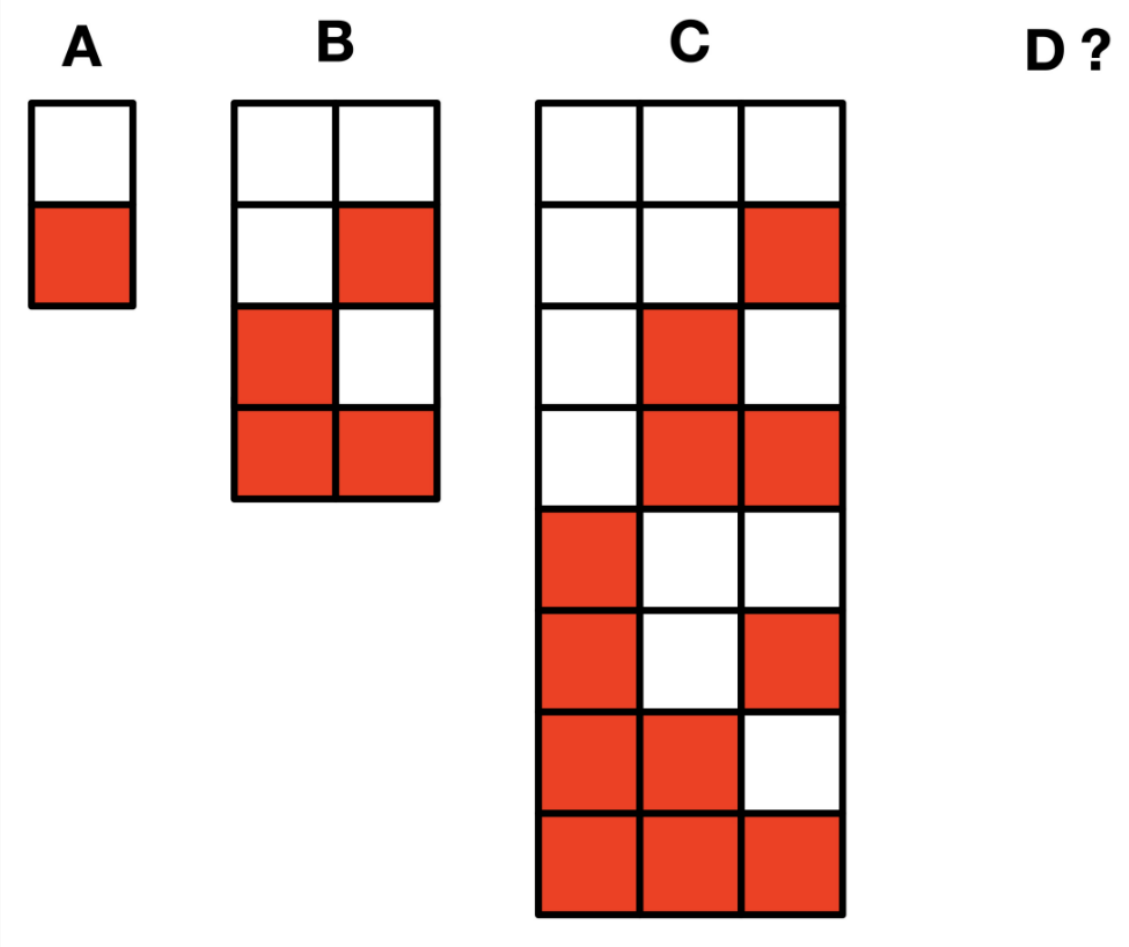
4 ?

Ans in next page

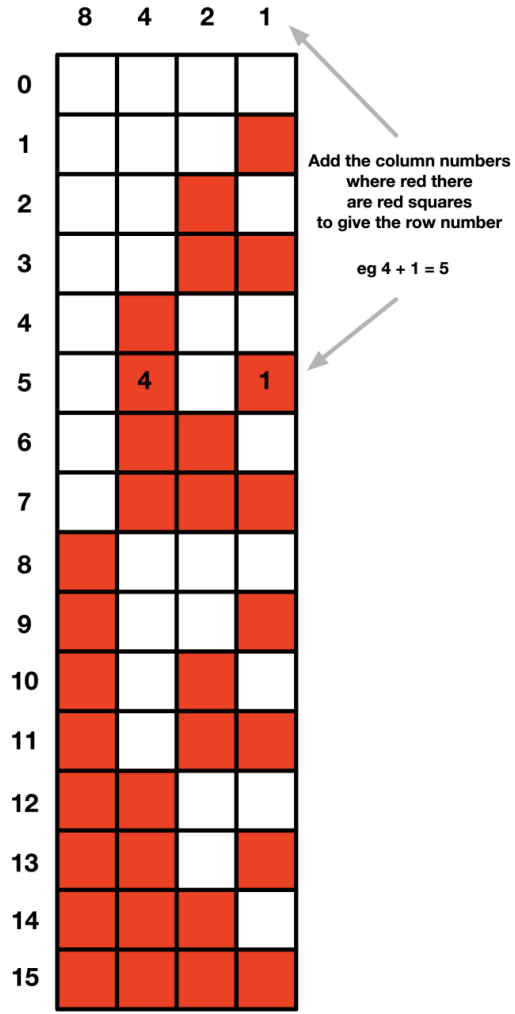
4



2. Find the pattern



Ans in next page


8
4
2
1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
4
1
Add the column numbers where red there are red squares to give the row number
eg 4 + 1 = 5

Why Pattern Recognition ?

- Patterns make our task simpler.
- Problems are easier to solve when they share patterns, because we can use the same problem-solving solution wherever the pattern exists.
- The more patterns we can find, the easier and quicker our overall task of problem solving will be.

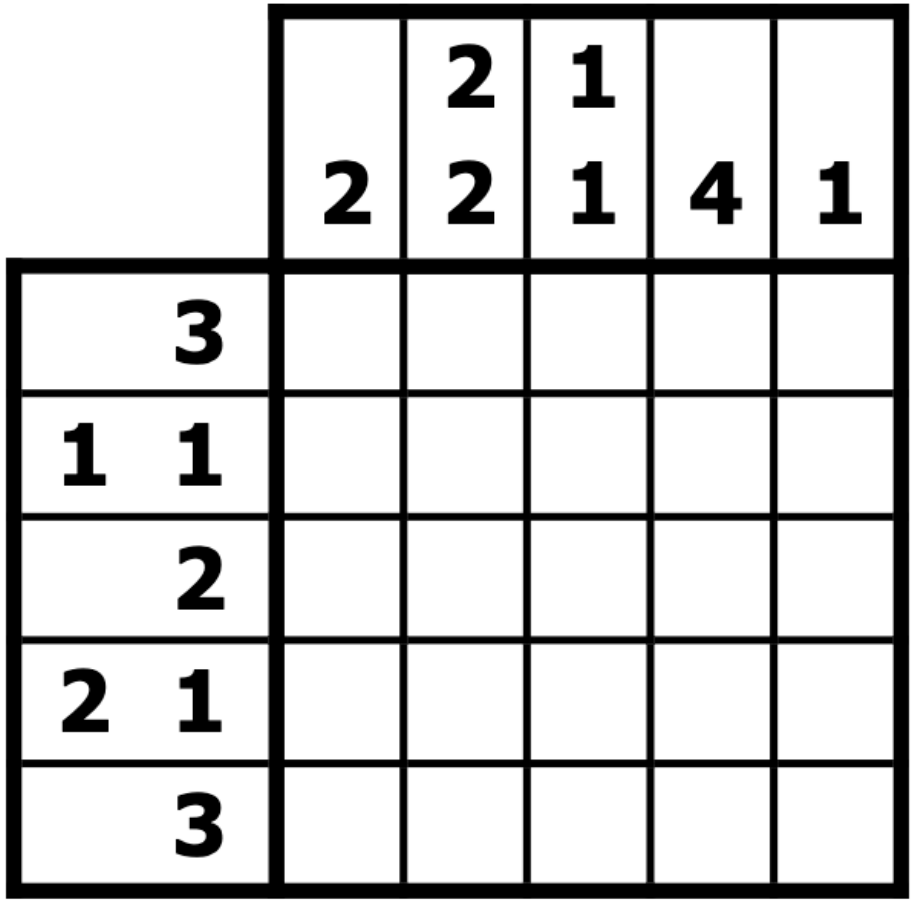
Nonograms

Nonograms is a logic puzzle with simple rules and challenging solutions.
The rules are

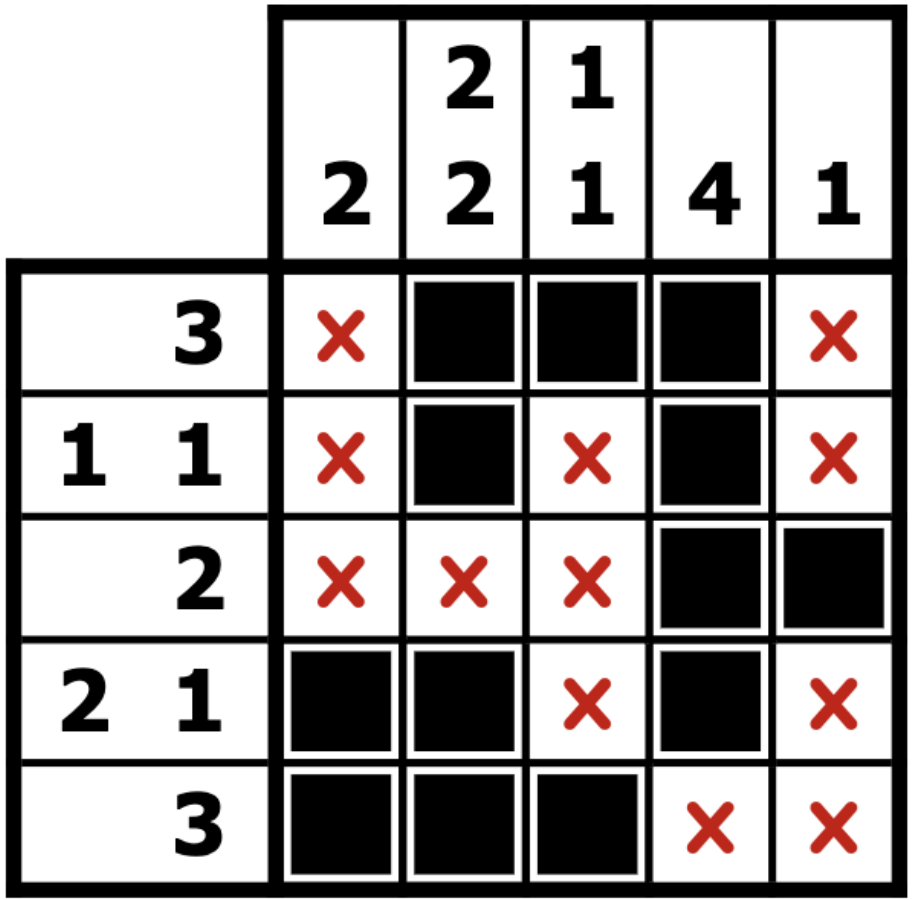
- You have a grid of squares, which must be either filled in black or marked with X.
- Beside each row of the grid are listed the lengths of the runs of black squares on that row.
- Above each column are listed the lengths of the runs of black squares in that column.
- Your aim is to find all black squares.

For more nonograms : Nonograms - online puzzle game

Puzzle 1 :



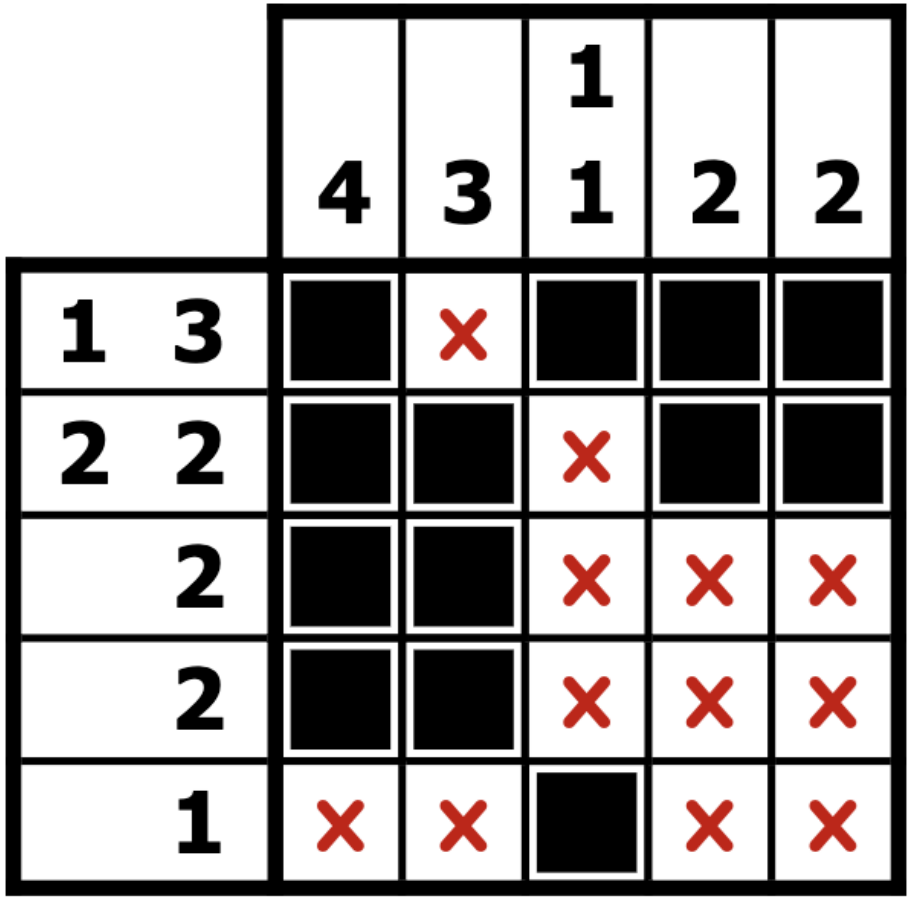
Ans in next page


2
2 2
1 1
4
1
3
1 1
2
2 1
3

Puzzle 2 :

		4	3	1 1	2	2
1	3					
2	2					
	2					
	2					
	1					

Ans in next page


4 3 1 1 2 2
1 3
2 2
2
2
1

Data Encryption

Definition: Encryption turns **readable data** (plaintext) into an **unreadable form** (ciphertext), so it cannot be understood until it is reversed back.

“The primary purpose of data encryption is to protect the confidentiality of digital information by turning readable data into an unreadable code that requires a specific key to unlock”

Note:

- Encryption is the general name for this process.
- Ciphering is the same action, just the word used when you're doing it to a specific piece of text, like a sentence.

Question : The grid in the next slide shows a substitution cipher, each letter of the alphabet (A–Z) has been replaced with a unique picture symbol.

Using the key, decode the two encoded lines

A	B	C	D	E	F	G	H	I	J	K	L	M
												
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
												

TEACHER:   

PUPIL:  



Ans in next page

Teacher: WHO SAYS OH, OH, OH?

Pupil: SANTA WALKING BACKWARDS

(The joke: Santa's "ho ho ho" backwards is "oh oh oh.")

Data Compression

Definition: Compression makes data smaller by removing repeated or unnecessary parts, so it takes up less space and can be reconstructed later.

Example :

Compression looks for repeated parts and stores them more briefly, without changing what the data means.

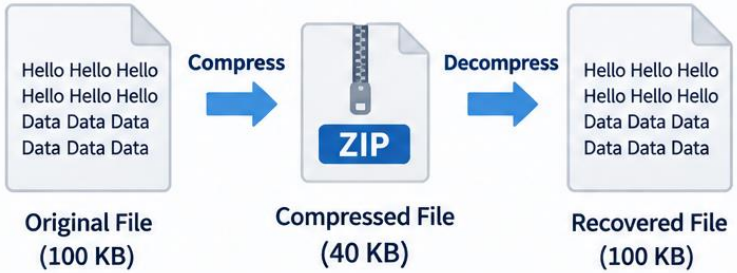
Original: **HAPPY HAPPY HAPPY**
Compressed: **HAPPY × 3**

Instead of writing the word three times, it's written once with a count. The full sentence can always be rebuilt from this shorter form, with no meaning lost.

Types of Data Compression

Lossless Compression

- The original data comes back **exactly** as it was.
- Used for text, code, and files where nothing can be lost.



Original File (100 KB) | Compressed File (40 KB) | Recovered File (100 KB)

Examples:

Text files | Source code | ZIP, RAR files

Lossy Compression

- The data comes back only **approximately**, but much smaller.
- Used for photos, music, and video.



Original Image (5 MB) | Compressed File (500 KB) | Recovered Image (≈ 5 MB) (slightly lower quality)

Examples:

Images (JPEG) | Audio (MP3) | Video (MP4)