Unit-IV

Visualization of groups: trees, graphs, clusters, Metaphorical visualization

Visualization of groups

Visualization of Groups is a branch of data visualization that focuses on how to represent **collections of related entities** — such as hierarchies, networks, clusters, or abstract categories — so that users can understand **relationships, structure, and similarities** within the data. Instead of focusing on **individual data points**, group visualization shows **how data points are organized into meaningful sets**.

•Many real-world datasets are **not flat tables**, but **structured into groups**:
•Hierarchies (family trees, organizational charts).
•Networks (social connections, IoMT devices).
•Clusters (customer groups, gene families).
•Abstract categories (topics, knowledge domains).
Group visualization helps users see:
•**Membership** → which item belongs to which group.
•**Relationships** → how groups connect.
•**Patterns** → similarities/differences across groups.

Main Types of Group Visualization

1.Trees (Hierarchies)

1. Show parent–child structures.
2. Examples: Org charts, file systems, taxonomies.

2.Graphs (Networks)

1. Show nodes and edges (entities + relationships).
2. Examples: Social networks, IoT networks, citation networks.

3.Clusters (Similarity Groups)

1. Show natural groupings of similar items.
2. Examples: Market segmentation, bioinformatics, topic modeling.

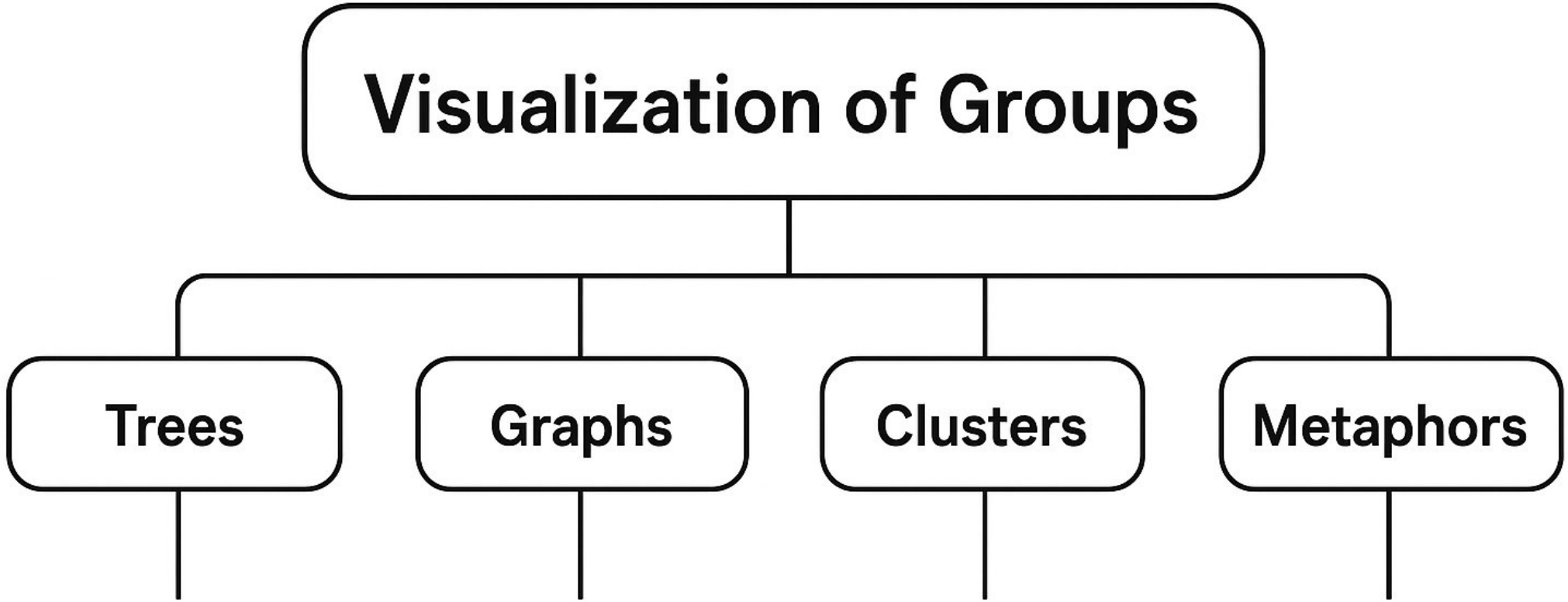
4.Metaphorical Visualizations

1. Use real-world metaphors (cities, galaxies, forests) to represent abstract group structures.

Example Applications

- **Healthcare:** Clustering patients by symptoms or IoMT sensor patterns.
- **Business:** Visualizing customer segments or product hierarchies.
- **Education:** Mapping topics into knowledge trees.
- **Biology:** Gene clusters, protein interaction networks, phylogenetic trees.

Visualization of Groups = techniques for representing **structured collections of entities** (hierarchies, relationships, clusters, metaphors) so people can **see patterns, connections, and group structures** in complex data.
It's especially useful when **data is relational or hierarchical**, not just numeric.


Visualization of Groups
Trees
Graphs
Clusters
Metaphors
• Org charts
• File systems
• Taxonomies
• Social networks
• IoT networks
• Citation networks
• Market segmentaion
• Bioinformatics
• Topic modeling
• Cities
• Galaxies
• Forests

Visualization of Groups – Trees

A **tree** is a **hierarchical structure** where:

- **Nodes** represent entities (objects, categories, or records).
- **Edges** represent **parent–child relationships**.
- There is a **root node** at the top, and multiple levels of descendants below it.

This makes trees ideal for showing **hierarchies, nested relationships, and categorical breakdowns**.

Examples

1.File System Directory Trees

1. Folders contain subfolders and files → classic tree.
2. Example: `C:\ → Documents → Projects → File1.txt`.

2.Organizational Charts

1. CEO → Managers → Employees.
2. Shows reporting structure.

3.Phylogenetic Trees (Biology)

1. Depict evolutionary relationships among species.
2. Root = common ancestor, branches = species divergence.

Visualization Techniques

1.Node-link Diagrams (Traditional Tree)

1. Each node connected by lines.
2. Good for small-to-medium hierarchies.
3. Example: Org charts.

2.Treemaps (Space-filling)

1. Represent hierarchical data as nested rectangles.
2. Size = value (e.g., sales volume), Color = attribute (e.g., profit).
3. Efficient for large datasets (used in Power BI, Tableau).

3.Radial Trees (Circular Layouts)

1. Nodes arranged in circles around the root.
2. Compact and aesthetically appealing.
3. Useful for visualizing large trees with many branches.

Advantages

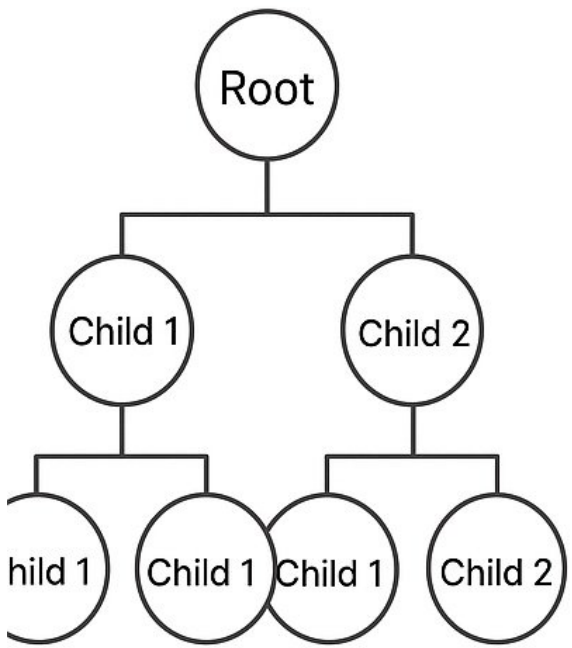
- Easy to **see hierarchies** and **nested relationships**.
- Treemaps and radial trees handle **large datasets efficiently**.

Challenges

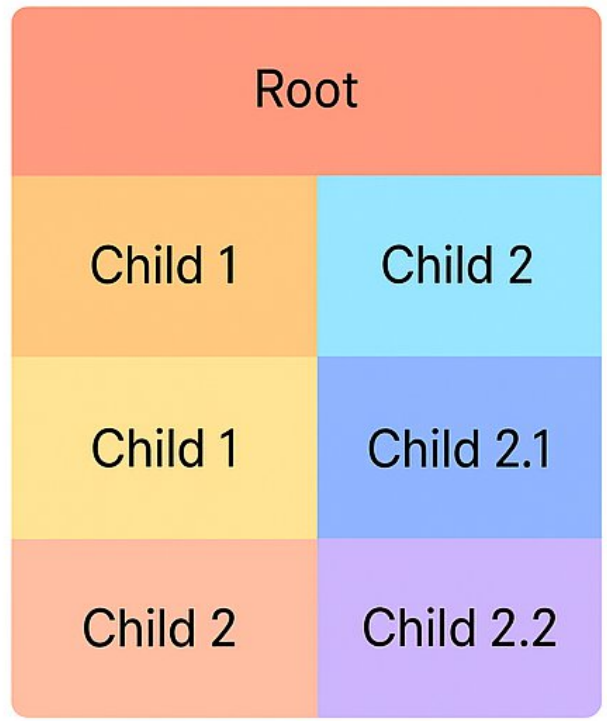
- Node-link trees get cluttered with large data.
- Treemaps may distort perception if rectangles are too small.

Trees

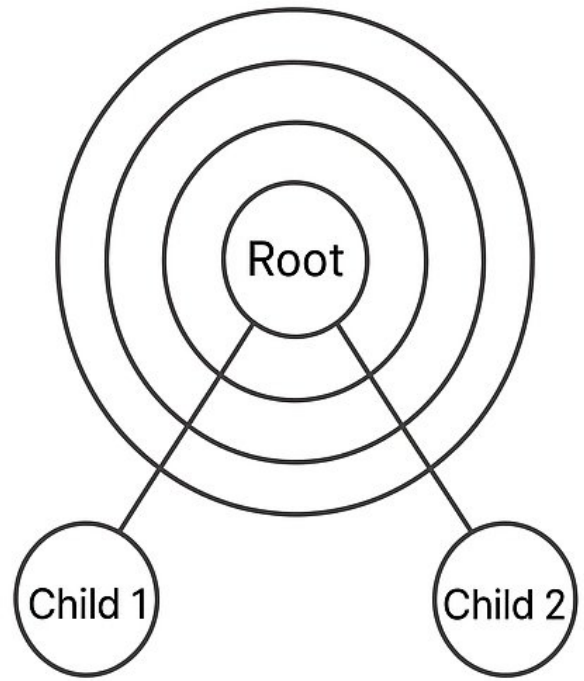
Node-link diagram



Treemap



Radial tree



Power BI Example: File System Directory Tree

Dataset Structure

We can represent a file system as a hierarchy with columns like:

Root	Folder	SubFolder	FileName	FileSize (KB)
C:\	Documents	Projects	Report.docx	120
C:\	Documents	Projects	Data.xlsx	300
C:\	Documents	Photos	Image1.png	450
C:\	Downloads	Software	Setup.exe	2048
D:\	Media	Music	Song.mp3	5000

2. Treemap Visualization

- **How:** Insert → **Treemap**.
- **Category fields:** Root → Folder → SubFolder → FileName.
- **Values field:** FileSize.
- **Result:**
  - Rectangles represent files.
  - Nesting shows hierarchy.
  - Size reflects file size.

👉 You can instantly see which folder/subfolder contains the largest files.

Decomposition Tree Visualization

- **How:** Insert → **Decomposition Tree**.
- **Measure:** `Sum(FileSize)`.
- **Dimensions:** Root → Folder → SubFolder → FileName.
- **Result:**
  - Start with total storage used.
  - Drill down interactively (Root → Folder → SubFolder → FileName).
  - Helps find “what’s eating space.”

4.Custom Org Chart/Hierarchy Visual

•From the Power BI marketplace → search **Hierarchy Chart** or **Tree Visual**.

•Map `Parent` and `Child` columns to show classic node-link diagrams (like Windows Explorer tree view).

Node	Parent	Size
Documents	C:\	420
Projects	Documents	300
Report.docx	Projects	120
Data.xlsx	Projects	180
Photos	Documents	100
Image1.png	Photos	100

This will give you a **true directory tree structure**.

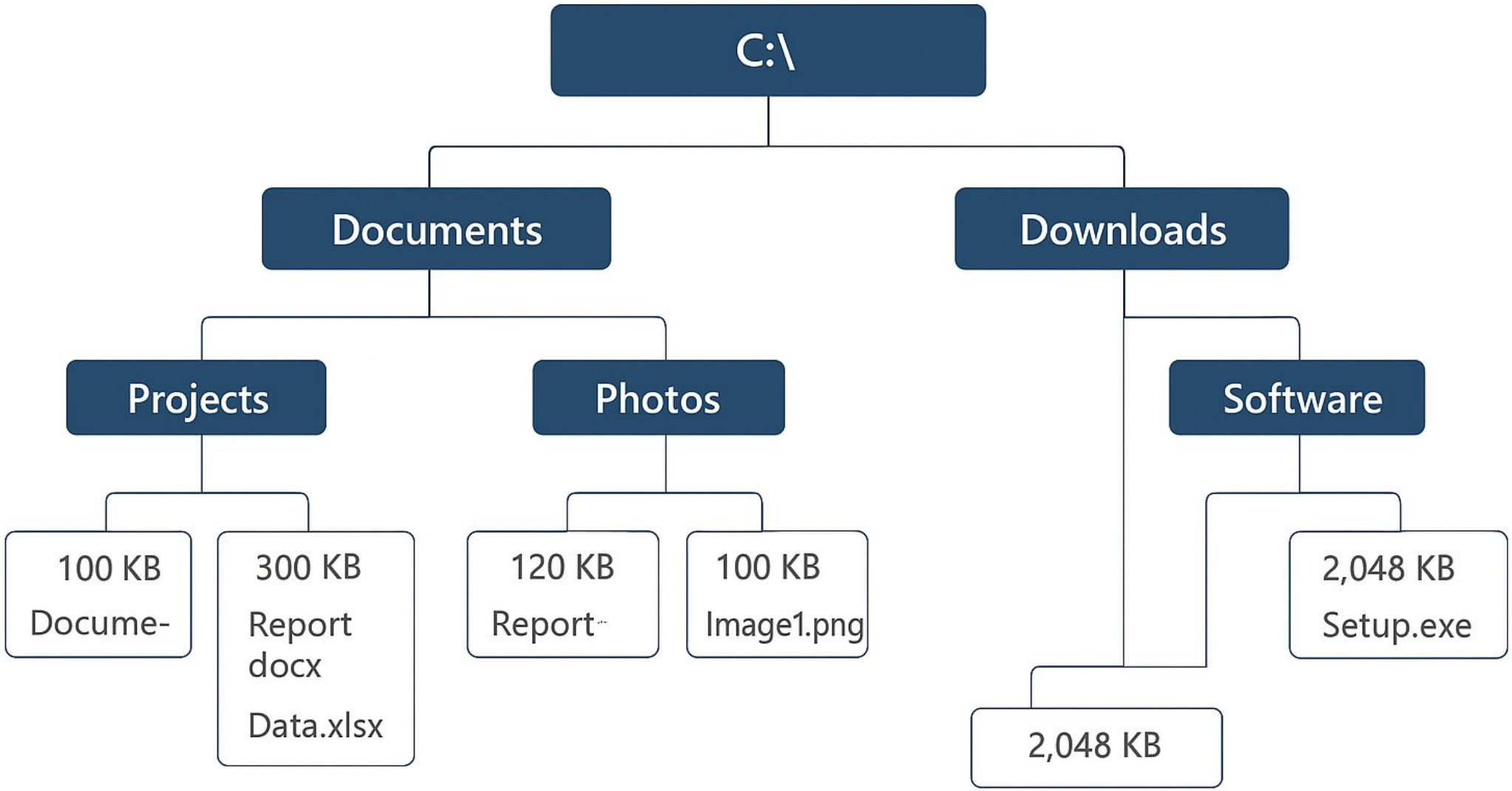
- **Treemap** → shows space usage visually.
- **Decomposition Tree** → shows interactive breakdown.
- **Hierarchy Chart (Custom Visual)** → shows real “file explorer” style trees.

How to Use in Power BI

1. Open **Power BI Desktop**.
2. Go to **Home → Get Data → CSV/Excel** and load the file.
3. Choose a visualization:
    1. **Treemap:** Root → Folder → SubFolder → FileName (Value = FileSize_KB).
    2. **Decomposition Tree:** Analyze total file size, drill down Root → Folder → SubFolder → FileName.
    3. **Hierarchy Chart (Custom Visual):** Parent–Child view of directories.

Decomposition Tree

Analyze total file size, then drill down



Radial Trees (Circular Layouts)

🔍 Definition

A **Radial Tree** is a variation of the traditional **node-link tree diagram**, but instead of extending vertically or horizontally, nodes are arranged in **concentric circles around a central root node**.

- **Root Node** → Center of the circle.
- **Children** → Placed in rings radiating outward.
- **Edges** → Drawn as straight or curved lines connecting parent and child.

Characteristics

- **Compact & Symmetric:** Uses space more efficiently than standard tree layouts.
- **Aesthetically Appealing:** Symmetry makes it good for presentation.
- **Scalable:** Can show a large number of nodes without becoming too long or wide.

Advantages

- Better space utilization compared to vertical/horizontal trees.
- Good for displaying **hierarchies with many branches**.
- Easy to spot balance vs. imbalance in hierarchies.

Challenges

- For very large datasets, edges may **overlap** and reduce readability.
- Linear hierarchy order may be harder to interpret compared to a classic tree.

Examples

1.Biology (Phylogenetic Trees):

1. Root = common ancestor.
2. Branches radiate out → different species.

2.Organization Structures:

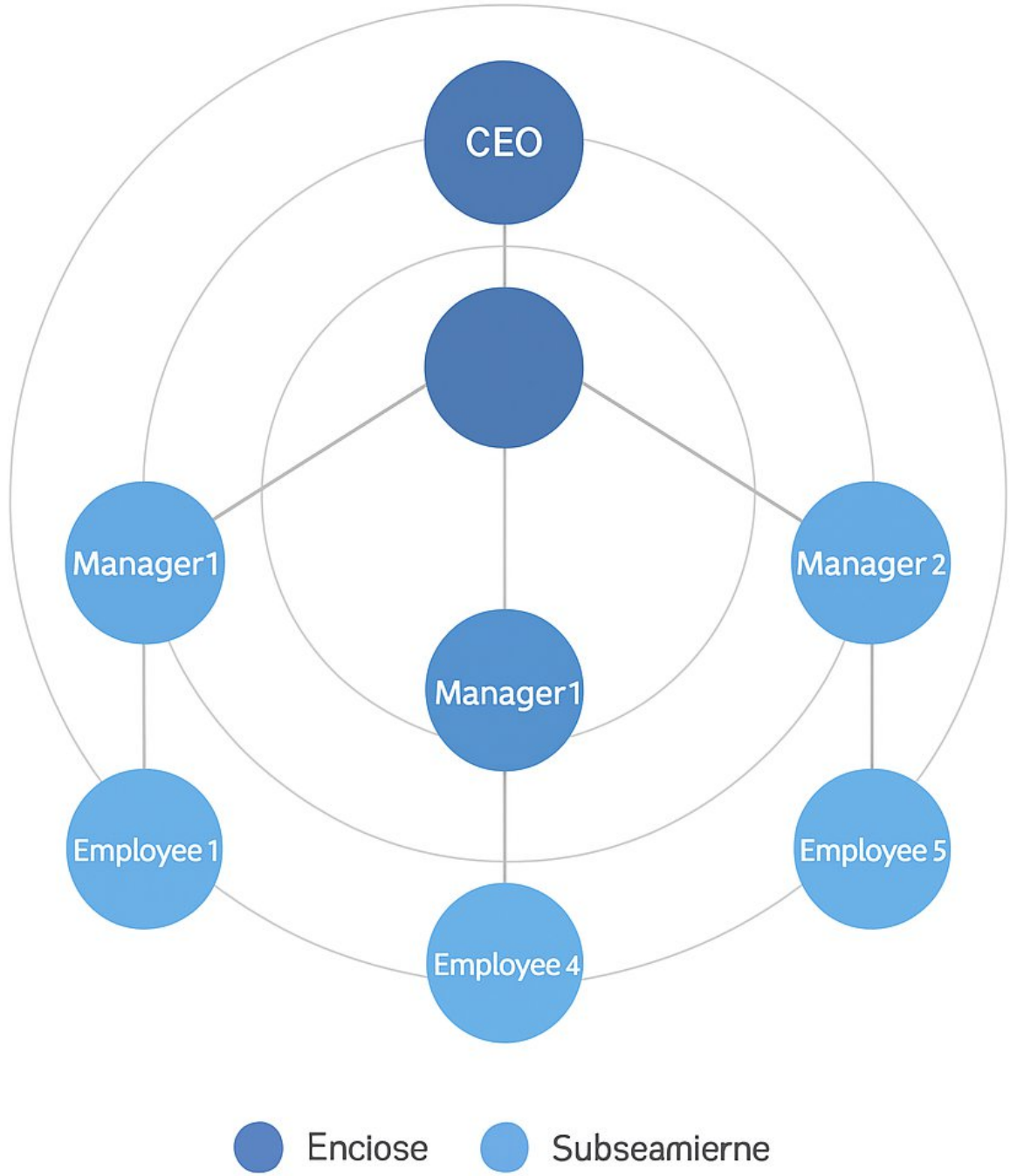
1. CEO at center, concentric circles for managers, employees, teams.

3.Knowledge Maps / Taxonomies:

1. Root = broad concept.
2. Branches = subfields/topics radiating outward.

Power BI Implementation

- Power BI **does not natively support radial trees**, but:
  - You can use **custom visuals** (Hierarchy Chart, Circle Packing, Sunburst).
  - **Sunburst Chart** (available in Power BI marketplace) is closest → concentric circles showing hierarchical levels.
  - For true Radial Trees → external tools (D3.js, R, Python) can be integrated into Power BI via **Python/R visuals**.


CEO
Manager1
Manager 2
Manager1
Employee 1
Employee 5
Employee 4
Enciose
Subseamierne

IoMT Patient Data Monitoring with Radial Tree

Healthcare Patient Data (IoMT Devices)

•**Use Case:** Monitoring patient data collected from IoMT (Internet of Medical Things).

•**Visualization:**

- Root = Patient.
- Ring 1 = Devices (ECG, Glucose Monitor, Oximeter).
- Ring 2 = Data Streams (Heart Rate, Blood Sugar, Oxygen).

•**Insight:** Doctors can see which device/patient contributes most to anomalies

PatientID	Category	SubCategory	Value
P001	Vitals	Heart Rate	80
P001	Vitals	Blood Pressure	120/80
P001	Vitals	Oxygen Level	98%
P001	Lifestyle	Activity	Active
P001	Lifestyle	Sleep Hours	7
P001	Medical	Medication	Yes
P001	Medical	Doctor Visit	Weekly

Load Data into Power BI

•Open **Power BI Desktop**

•Go to **Home → Get Data → Excel/CSV** (or SQL, API, IoT Hub)

•Load your IoMT dataset.

3. Use Hierarchical Visualization

Power BI doesn't have a native radial tree, but you can use **custom visuals** from the AppSource:

1.Go to **Visualizations → Get more visuals**

2.Search for:

1. **Hierarchy Chart by Akvelon**
2. **Sunburst Chart**
3. **TreeViz**
4. **Network Navigator**

For **radial tree**, "Hierarchy Chart by Akvelon" or "TreeViz" works best.

4. Build the Radial Tree

•Drag **Category** → parent node

•Drag **SubCategory** → child node

•Drag **PatientID** as the root

•Drag **Value** into details/tooltip

This creates a radial structure showing each patient's data across vitals, lifestyle, and medical records.

5. Enhance Dashboard

•**Conditional Formatting**

- Red for abnormal vitals
- Green for healthy values

•**Filters / Slicers**

- PatientID
- Date range
- Device type (if IoT devices are included)

•**Tooltip Drillthrough**

- Clicking on "Heart Rate" can open a time-series line chart

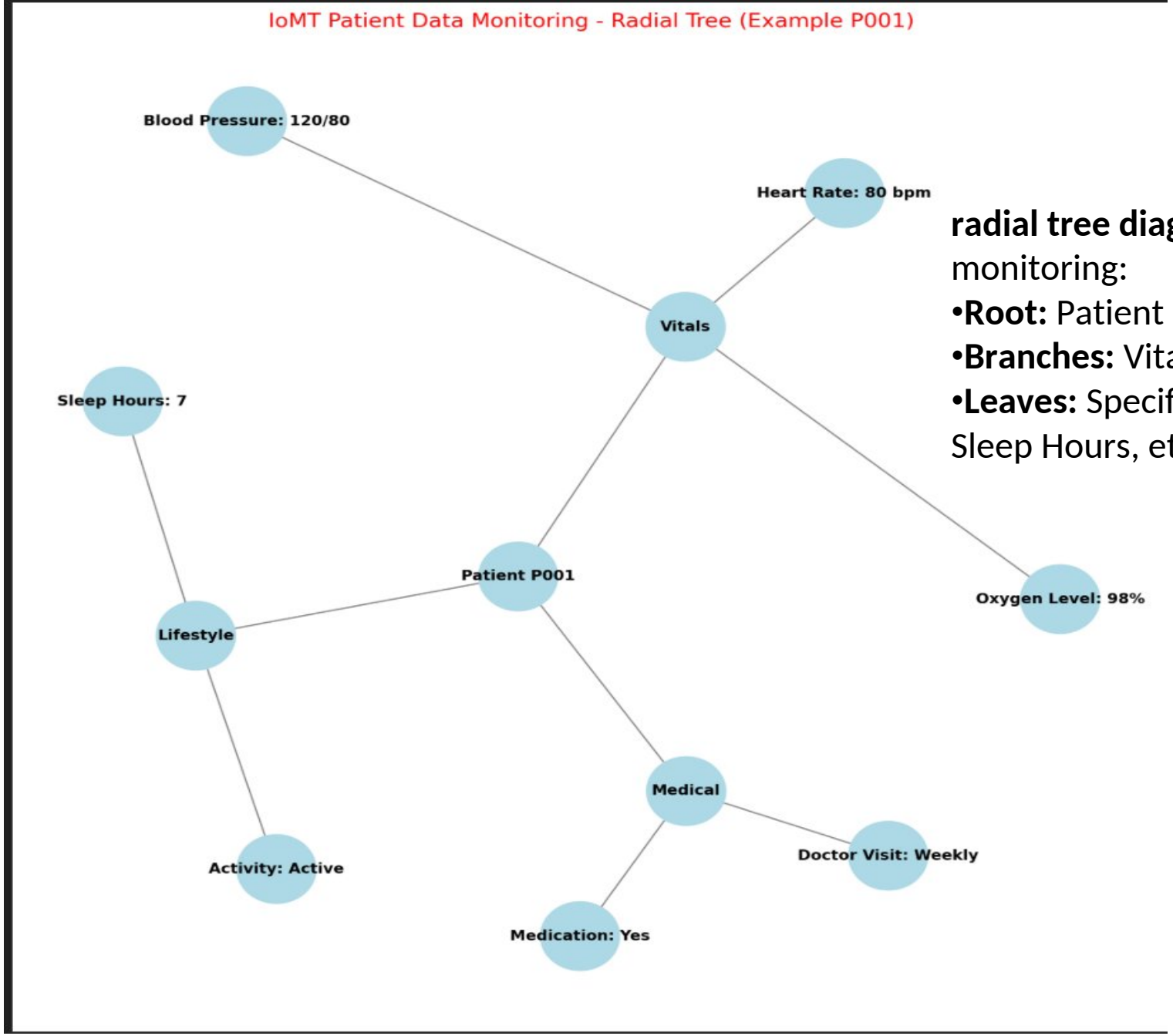
6. Optional: Real-Time IoMT Monitoring

If data comes from IoT devices:

•Use **Power BI Streaming Dataset**

•Connect with **Azure IoT Hub / Stream Analytics**

•Radial tree auto-updates with patient vitals.



radial tree diagram for one patient (**P001**) showing IoMT monitoring:

- **Root:** Patient
- **Branches:** Vitals, Lifestyle, Medical
- **Leaves:** Specific data (Heart Rate, Blood Pressure, Activity, Sleep Hours, etc.)

Graphs

Graphs are a fundamental group of data visualizations used to represent **relationships between entities**.

1. Definition

•A **graph** consists of:

- **Nodes (vertices):** Entities or objects (e.g., people, computers, proteins).
- **Edges (links):** Relationships or interactions between nodes (e.g., friendships, connections, flows).

2. Types of Graphs in Visualization

1.Node-Link Diagrams

1. Direct visual mapping of nodes and edges.
2. Example: Social network, citation network.

2.Adjacency Matrix

1. Represents relationships in tabular form (rows & columns).
2. Good for dense graphs.
3. Example: Flight connections between cities.

3.Hierarchical Graphs

1. Trees, dendrograms, radial trees.
2. Example: Organization chart, file system.

4. Clustered / Community Graphs

1. Detecting groups within a graph (community detection).
2. Example: Online community networks.

5.Specialized Graph Layouts

1. Force-directed layouts (natural clustering).
2. Circular or radial layouts (flows, balance).
3. Bipartite graphs (two distinct sets of nodes).

. Applications of Graph Visualization

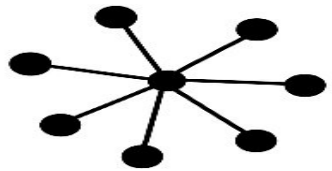
- **Social networks** → friendships, followers, influence.
- **Biological networks** → genes, proteins, metabolic pathways.
- **Computer networks** → internet topology, communication.
- **Recommendation systems** → users ↔ products.
- **Knowledge graphs** → semantic web, linked data.

Data Visualization Groups – Graphs

Type of Graph Visualization	Characteristics	Example
Node-Link Diagram	Nodes shown as points, edges as lines. Good for sparse networks.	Social network (friends, followers).
Adjacency Matrix	Rows & columns = nodes, cell = connection. Better for dense graphs.	Flight connections between cities.
Hierarchical Graph (Tree, Dendrogram)	Parent-child relationships, shown as tree or radial layout.	Organization chart, family tree.
Clustered / Community Graph	Groups of nodes with stronger internal connections.	Online community groups in Twitter.
Force-Directed Layout	Physics-based layout, nodes repel and edges attract. Natural grouping.	Citation network visualization.
Circular / Radial Layout (Chord Diagram)	Nodes arranged in a circle, edges/arcs show relationships.	Trade flows between countries.
Bipartite Graph	Two distinct sets of nodes, edges connect only across sets.	Users ↔ Movies in recommendation system.

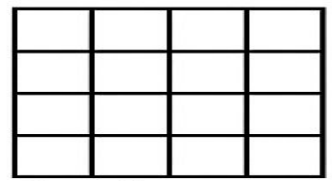
Data Visualization Groups – Graphs

Node-Link Diagram



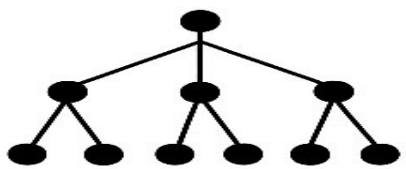
Nodes shown as points, edges as lines. Good for sparse networks.

Adjacency Matrix



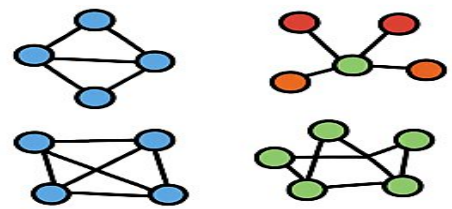
Rows & columns = nodes, cell = connection. Better for dense graphs

Hierarchical Graph (Tree, Dendrogram)



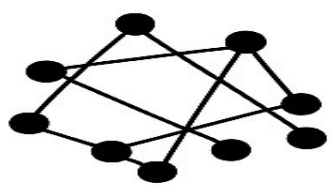
Parent-child relationships, shown as tree or radial layout

Clustered / Community Grapph



Groups of nodes with stronger internal connections

Force-Directed Layout



Physics-based layout, nodes repel and edges attract. Natural grouping

Circular / Radial Layout (Chord Diagram)



Two distinct sets of nodes, edges connect only across :

Biological networks → genes, proteins, metabolic pathways.

Types of Biological Networks

1.Gene Regulatory Networks (GRNs)

1. Nodes = genes, edges = regulatory interactions.
2. Shows how genes activate or repress each other.

2.Protein–Protein Interaction Networks (PPIs)

1. Nodes = proteins, edges = interactions (binding, signaling).
2. Useful for drug discovery and disease analysis.

3.Metabolic Pathways

1. Nodes = metabolites/enzymes, edges = biochemical reactions.
2. Example: Glycolysis pathway.

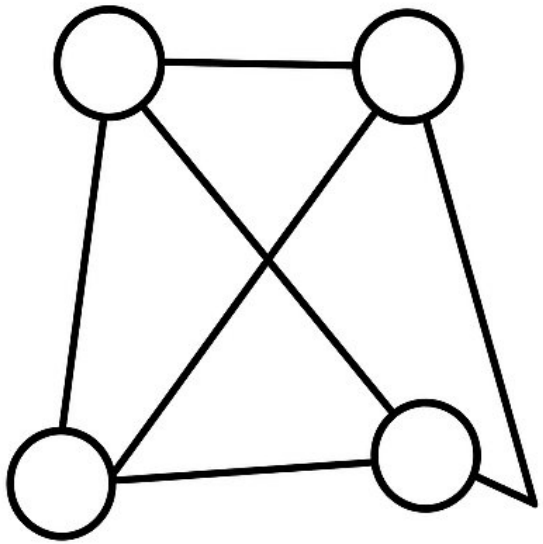
4.Neural Networks (Biological, not AI)

1. Nodes = neurons, edges = synaptic connections.
2. Example: Brain connectivity maps.

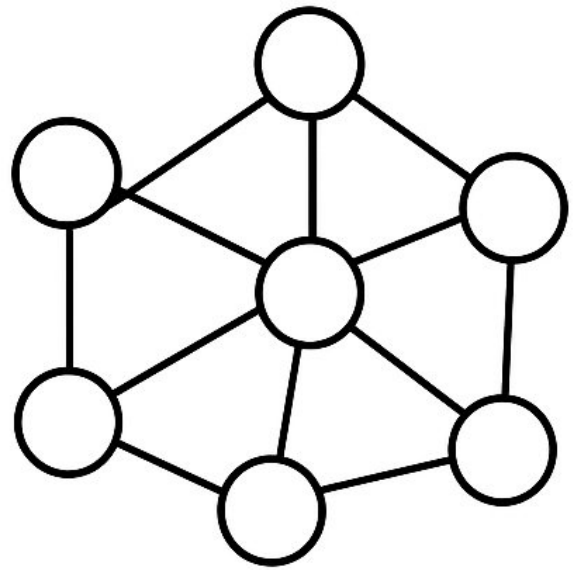
Why Use Graph Visualization in Biology?

- To **understand system behavior** (how molecules interact).
- To **identify clusters/modules** (e.g., groups of genes related to a disease).
- To **predict new drug targets** by analyzing hub nodes.
- To **visualize pathways** for teaching and research.

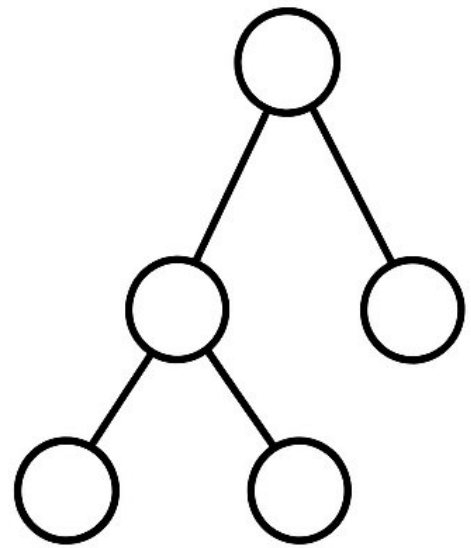
Biological Networks



Gene Regulatory Network



Protein-Protein Interaction Network



Metabolic Pathway

Gene Regulatory Networks (GRNs)

. Genes Table (Nodes)

This table stores the list of genes and their attributes.

GeneID	GeneName	Type	Function
G1	Gene A	Activator	Transcription factor
G2	Gene B	Repressor	Cell cycle control
G3	Gene C	Activator	Stress response
G4	Gene D	Repressor	DNA repair
G5	Gene E	Activator	Metabolism
G6	Gene F	Repressor	Apoptosis
G7	Gene G	Activator	Signal transduction
G8	Gene H	Repressor	Immune response
G9	Gene I	Activator	Protein synthesis
G10	Gene J	Repressor	Growth regulation

Interactions Table (Edges)
This table stores the relationships between genes (who regulates whom, and how).

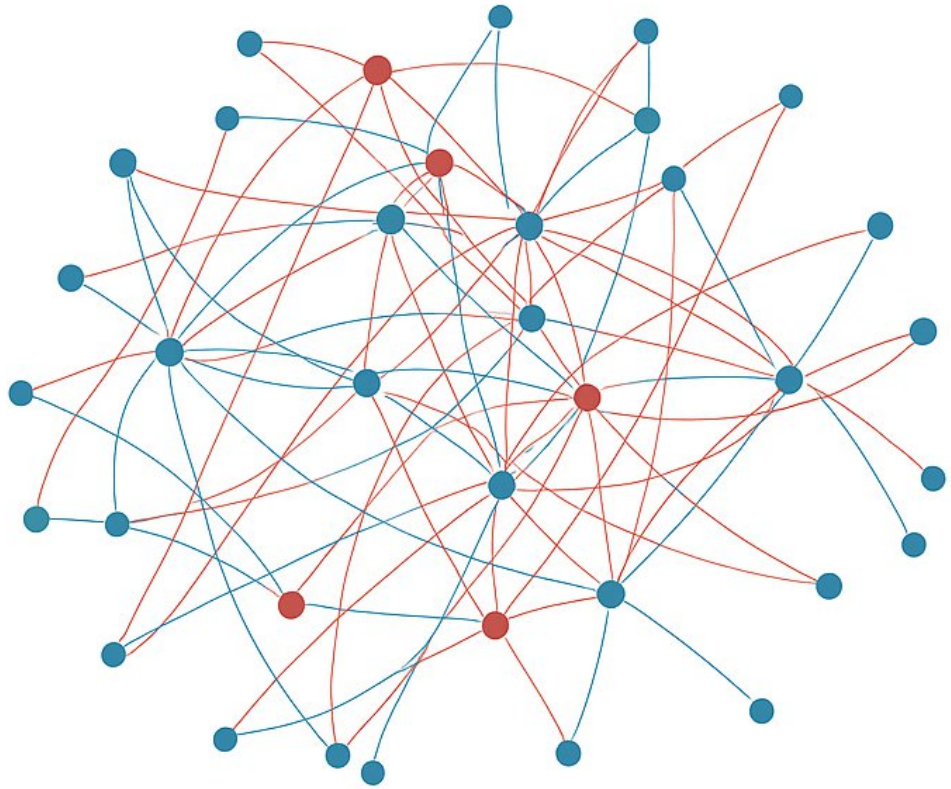
InteractionID	SourceGeneID	TargetGeneID	InteractionType	Strength
1	G1	G2	Activation	0.8
2	G2	G3	Repression	0.6
3	G1	G4	Activation	0.9
4	G3	G5	Activation	0.7
5	G4	G6	Repression	0.5
6	G5	G7	Activation	0.8
7	G6	G8	Repression	0.6
8	G7	G9	Activation	0.9
9	G8	G10	Repression	0.7
10	G9	G1	Activation	0.6

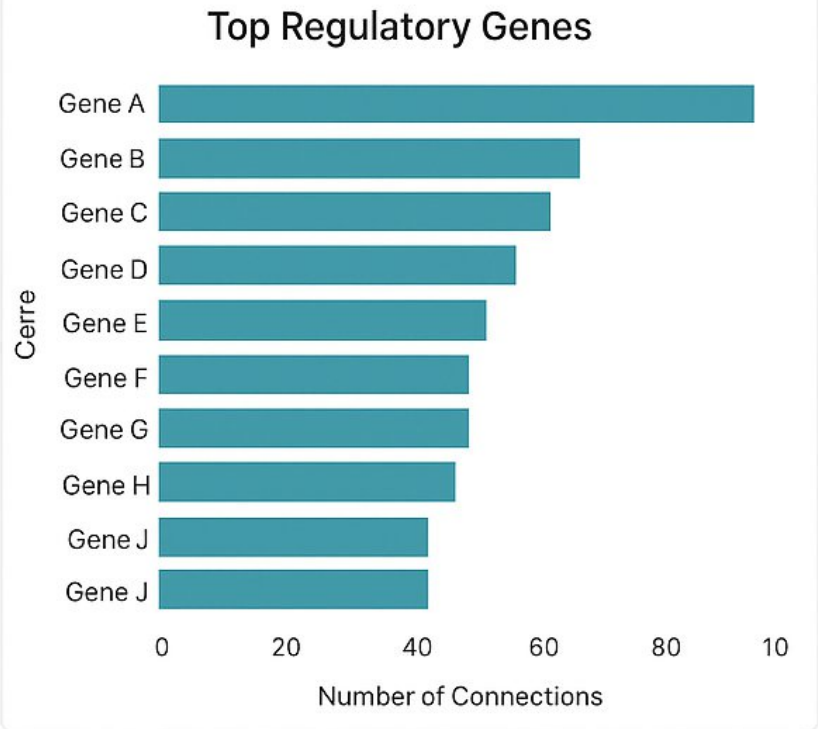
3. Regulatory Changes Table (Optional – For Time Series)
This table tracks changes in regulatory interactions across dates (used for the line chart)

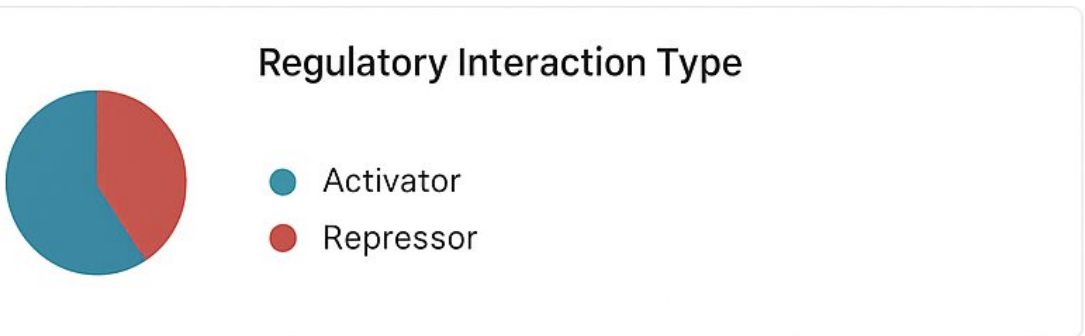
Date	NumberOfChanges
2025-01-04	15
2025-01-05	28
2025-01-06	20
2025-01-07	35

With these three tables, you can:
•Build the **Network Graph** from `Genes` + `Interactions`.
•Build the **Top Regulatory Genes (Bar Chart)** by counting `SourceGeneID` occurrences.
•Build the **Pie Chart** from `InteractionType`.
•Build the **Line Chart** from `Regulatory Changes`.

Gene Regulatory Networks (GRNs)









Clusters

Data points are grouped based on similarity. In visualization, clusters are used to show natural groupings, patterns, or communities within a dataset.

Types of Cluster Visualization

1.Scatterplots with Clusters

1. Different clusters are shown with distinct colors or shapes.
2. Example: Customer segmentation in marketing.

2.Hierarchical Clustering (Dendrograms)

1. A tree-like structure that shows how data points group together at different similarity thresholds.

3.Heatmaps with Clustering

1. Often used in bioinformatics (e.g., gene expression data).
2. Rows and columns are reordered to group similar values.

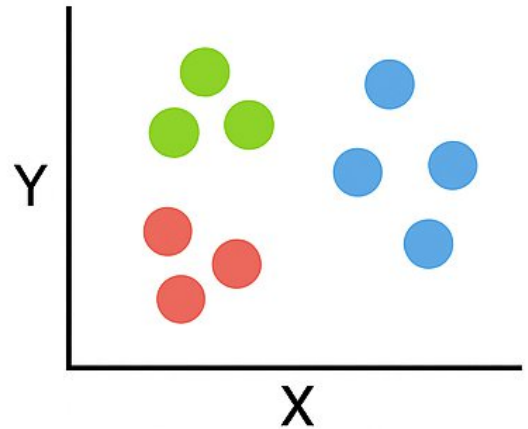
4.Network Clustering

1. Groups are detected in graph networks (e.g., communities in social networks).

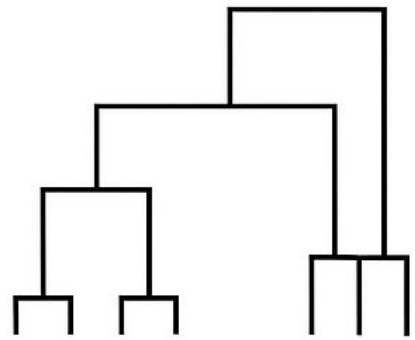
5.Clustered Bar/Column Charts

1. Used when categorical data is grouped into clusters.

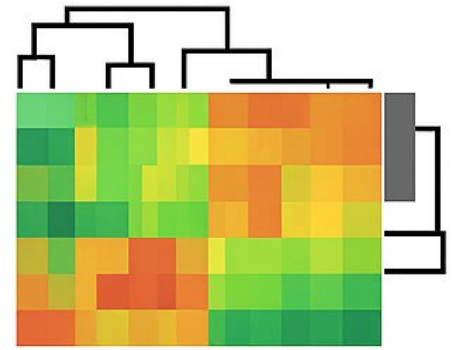
Types of Clusters



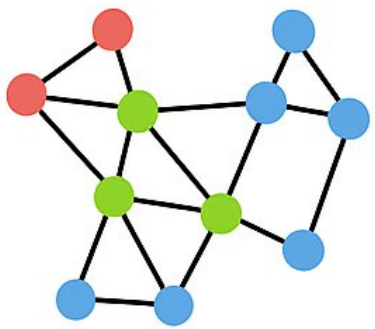
Scatterplot with Clusters



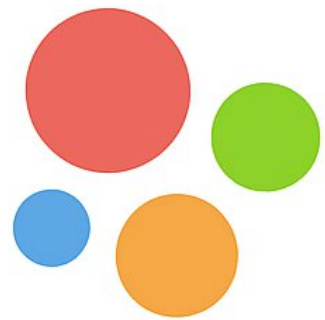
Hierarchical Clustering (Dendrogram)



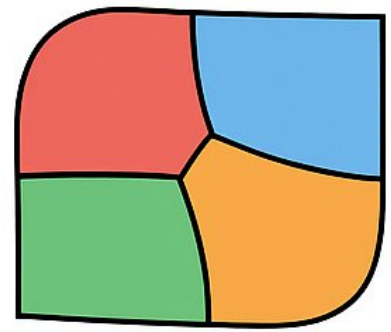
Heatmap with Clustered Rows/Clumns



Clustered Netwk Graph



Cluster Bubble Chart



Voronoi / Cluster Regions

Ex: Customer Segmentation with Hierarchical Clustering

Step 1: Sample Dataset
Suppose we have customer data with two features: **Annual Income** and **Spending Score**.

CustomerID	AnnualIncome	SpendingScore
1	25	35
2	28	40
3	30	38
4	65	80
5	70	78
6	72	82
7	48	20
8	50	22
9	52	18

Step 2: Perform Hierarchical Clustering (outside Power BI)

Since Power BI does not directly perform clustering with dendrograms:

•Do clustering in **Python or R** (using scikit-learn or scipy).

•Export results (cluster labels for each customer).

For example, using Python:

```
from scipy.cluster.hierarchy import linkage, fcluster
import pandas as pd

# Data
data = pd.DataFrame({
    'AnnualIncome': [25,28,30,65,70,72,48,50,52],
    'SpendingScore': [35,40,38,80,78,82,20,22,18]
})

# Hierarchical Clustering
Z = linkage(data, method='ward')  # Ward’s method
clusters = fcluster(Z, 3, criterion='maxclust')  # 3 clusters

data['Cluster'] = clusters
print(data)
```

Step 3: Load into Power BI

Now load the dataset into Power BI with the new **Cluster column**.

Example table after clustering:

CustomerID	AnnualIncome	SpendingScore	Cluster
1	25	35	1
2	28	40	1
3	30	38	1
4	65	80	2
5	70	78	2
6	72	82	2
7	48	20	3
8	50	22	3
9	52	18	3

Step 4: Power BI Visualizations

1.Scatterplot

1. X-axis → `AnnualIncome`
2. Y-axis → `SpendingScore`
3. Legend → `Cluster`
4. Each cluster will be color-coded.

2.Cluster Sizes (Bar Chart)

1. X-axis → `Cluster`
2. Y-axis → Count of `CustomerID`

3.Average Spending per Cluster (Card or Table)

1. Use DAX:

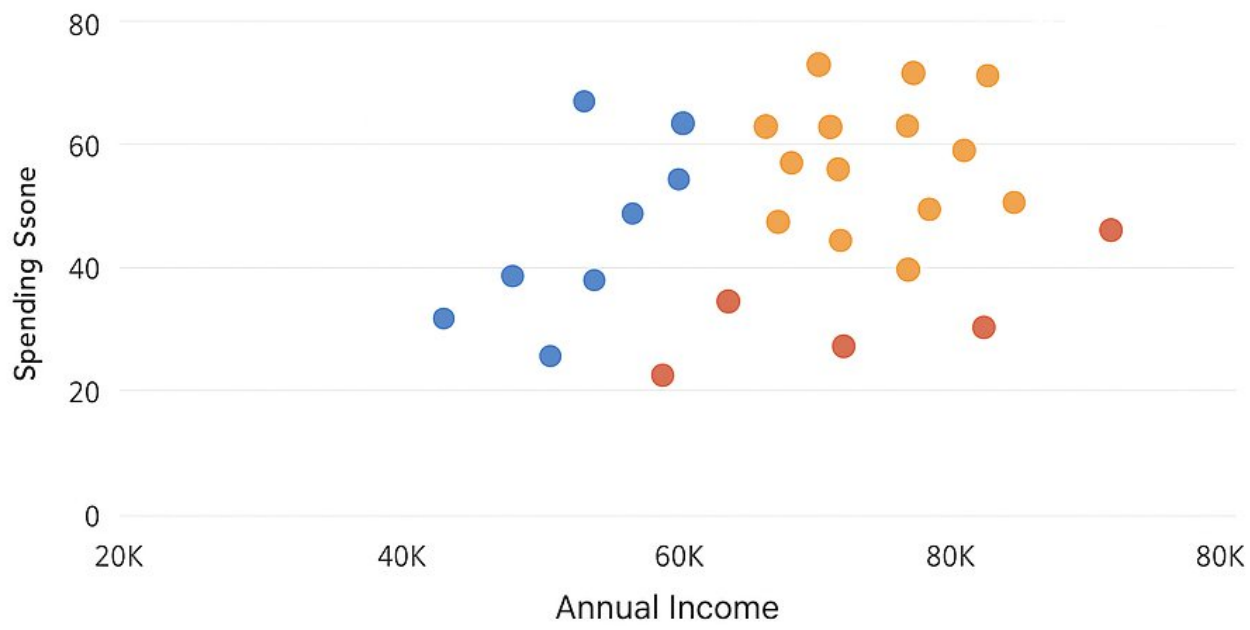
```
AvgSpending = AVERAGE(Data[SpendingScore])
```

Optional – Dendrogram

•Power BI doesn't have a built-in dendrogram, but you can use a **custom R or Python visual** to plot it.

Customer Segmentation Dashboard (Hierarchical Clustering)

Clusters by Income vs Spending Score



Cluster Size



Total Customers

9

Clusters

3

Highest Spending Cluster

Cluster Profile

Cluster	Avg Income	Avg Spending Score	Count of Customers
Cluster 1	27.7	37.7	3
Cluster 2	69.0K	80.0	2
Cluster 3	50.0K	20.0	4

Metaphorical visualization

In data visualization, **metaphorical visualization** means representing data using familiar real-world objects, symbols, or metaphors instead of abstract shapes (bars, lines, etc.). This makes it easier to understand patterns and relationships intuitively.

Examples of Metaphorical Visualization:

- **Tree metaphor** → used for **hierarchies** (e.g., family trees, organizational structures).
- **Road/Path metaphor** → used for **processes or timelines**.
- **Galaxy/Constellation metaphor** → used for **networks and clusters**.
- **Building metaphor** → used for **modular data systems**.

