Unit-III

Classification of visualization systems: Types of visualization systems: scientific vs. information visualization, Interaction techniques: filtering, zooming, brushing, linking,Misleading visualizations – how to avoid bias and distortion,Visualization of 1D (temporal), 2D (maps, scatterplots), and text data

Types of Visualization Systems

1. Scientific Visualization

- Focuses on **data from physical or natural phenomena**.
- Used in domains like physics, biology, engineering, medicine.
- Examples:
  - Medical imaging (CT scans, MRI)
  - Weather simulation (wind patterns, rainfall maps)
  - Molecular structures

- **Scientific Visualization = Physics of the real world** (continuous data).
- **Information Visualization = Patterns in abstract data** (categorical/relational).

2. Information Visualization

- Deals with **abstract, non-physical data**.
- Goal: Make sense of large, complex information.
- Examples:
  - Text visualization (word clouds, topic maps)
  - Financial dashboards (stock trends, revenue flows)
  - Network diagrams (social networks, web links)

3. Interaction Techniques in Visualization Systems

- **Filtering** → Selecting subsets of data.
- **Zooming** → Viewing data at different levels of detail.
- **Brushing & Linking** → Highlighting data in one view updates related views.

4. Data Types Visualization

- **1D Data (Temporal)** → Line charts, timelines.
- **2D Data (Spatial/Maps)** → Scatterplots, choropleth maps.
- **Text Data** → Word clouds, topic clusters.

Scientific vs. Information Visualization

- **1. Scientific Visualization**

Definition

•The use of computer graphics to represent **data derived from physical or natural phenomena**.

•The data is usually numerical, continuous, and often results from simulations, sensors, or experiments.

Characteristics

•Deals with **real-world, physical measurements**.

•Data is often **multi-dimensional** (3D, 4D, spatiotemporal).

•Visualization techniques are highly domain-specific.

•Aims to **simulate, analyze, and predict physical processes**.

Examples

•Weather and climate simulations (temperature, rainfall, wind flow).

•Medical imaging (MRI, CT scans, X-rays).

•Molecular and chemical structures.

•Engineering simulations (stress analysis, fluid dynamics, aerodynamics).

Example: Weather Forecasting using Scientific Visualization

- **Step 1: Data Source**

•Meteorological departments collect **numerical, physical data**:

- Temperature
- Humidity
- Air pressure
- Wind speed and direction
- Rainfall

- **Step 2: Data Nature**

•This data is **continuous** (varies over space and time).

•It comes from **sensors, satellites, and climate simulations**.

Step 3: Visualization Techniques

•The raw numbers are transformed into **visual representations** such as:

- **Heat maps** for temperature variations.
- **Vector field plots** for wind direction and speed.
- **3D models** for storm/cloud movement.
- **Animated time series** for rainfall prediction.

◆ **Step 4: Outcome**

•Scientists and meteorologists can:

- Predict **storms, cyclones, and rainfall**.
- Identify **climate patterns** (El Niño, La Niña).
- Help governments issue **weather warnings**.

Power BI Dashboard – Weather Forecasting Example

Dataset Fields

Your weather dataset could include:

- **Date & Time**
- **Location (City, Region, Country, Latitude, Longitude)**
- **Temperature (°C/°F)**
- **Humidity (%)**
- **Rainfall (mm)**
- **Wind Speed (km/h)**
- **Wind Direction**
- **Air Pressure (hPa)**

Step 1: Import Data

1. Collect historical or simulated weather data.
2. In **Power BI Desktop** → *Get Data* → Excel/CSV → Import dataset.
3. Check data types (Date = Date/Time, Temp = Decimal, Wind = Number, etc.).

Date	Location	Temperature (°C)	Humidity (%)	Rainfall (mm)	Wind Speed (km/h)	Wind Direction	Air Pressure (hPa)
2021-01-15	China	31	66	0.7	23	SW	981
2021-01-26	China	37.8	41	50.5	21	SW	1034
2021-01-25	China	33.4	67	10	33	NW	1024
2021-01-28	China	12.2	61	54	40	SE	993
2021-01-13	China	37.1	74	6.1	33	SE	985
2021-01-19	India	31.5	86	4.9	36	E	1044
2021-01-17	India	12.8	55	2.7	19	E	1021
2021-01-17	China	21.6	68	59	26	S	993
2021-01-06	India	35.5	40	26.2	29	NE	993
2021-01-02	USA	27.2	47	58.7	23	NE	1030
2021-01-30	USA	18.9	54	23.8	23	NW	1027
2021-01-17	USA	13.5	60	40.4	24	NW	980
2021-01-08	UK	34.3	76	15.4	25	SE	1022
2021-01-21	China	11.5	83	16.8	31	NW	1021
2021-01-26	India	15.6	45	22.1	5	W	992
2021-01-01	USA	17.8	66	9.6	16	SE	1021
2021-01-24	China	23.2	90	29.1	27	N	1029
2021-01-30	UK	31.1	69	5.7	11	NE	1005
2021-01-08	USA	26.6	77	4	10	NW	1040

Step 2: Create Measures (DAX)

```
Avg Temperature = AVERAGE(Weather[Temperature])
Avg Humidity = AVERAGE(Weather[Humidity])
Total Rainfall = SUM(Weather[Rainfall])
Max Wind Speed = MAX(Weather[Wind Speed])
```

Step 3: Build Visuals

1.KPI Cards

1. Average Temperature
2. Average Humidity
3. Total Rainfall
4. Max Wind Speed

2.Charts

1. **Line Chart** → Temperature Trend over Time (daily/hourly).
2. **Column Chart** → Monthly Rainfall totals.
3. **Scatter Plot** → Temperature vs. Humidity.

3.Map Visual (Filled Map)

1. Display temperature, rainfall, or wind speed across regions.

4.Wind Direction Visualization

1. Use a **Radar/Polar Chart (custom visual)** to show wind direction distribution.

5.Slicers

1. Date (Year, Month, Day).
2. Location (Country, Region, City).

Step 4: Dashboard Layout

- **Top row** → KPI Cards (Temp, Humidity, Rainfall, Wind).
- **Left side** → Slicers (Date, Region).
- **Center** → Line Chart (Temp trend) & Column Chart (Rainfall).
- **Right side** → Map (Weather by region) + Wind Direction chart.

Step 5: Insights from Dashboard

- Identify **temperature trends** (daily, seasonal).
- Track **rainfall patterns** across regions.
- Detect **humidity vs temperature correlation**.
- Monitor **wind speed & direction** for storm predictions.
- Compare weather conditions across multiple locations.

Weather

Avg Temperature	Avg Humidity	Total Rainfall	Max Wind Speed
21.5	68%	123 mm	35

Date

- ☐ 2020
- ☐ 2021
- ☐ 2022

Location

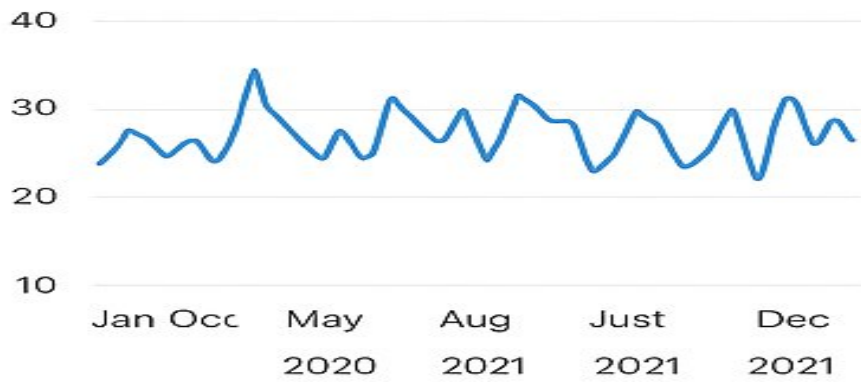
- ☐ USA
- ☐ India
- ☐ UK
- ☐ China

$ 20

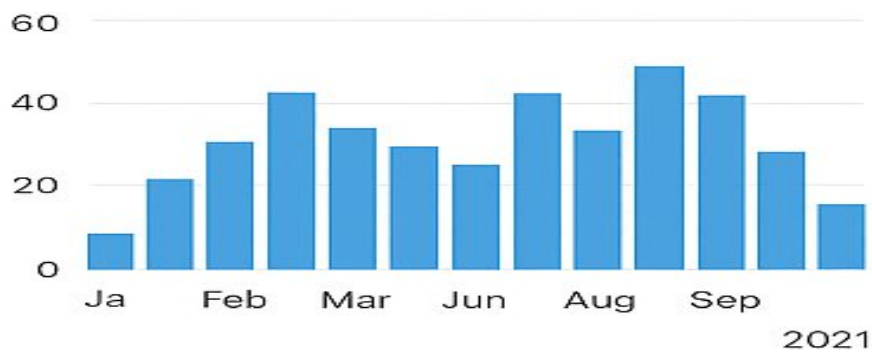
Yel a Or | mi re

29

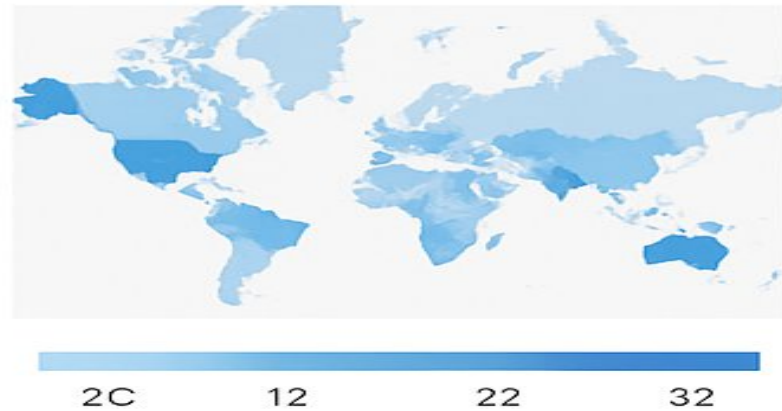
Temperature Trend



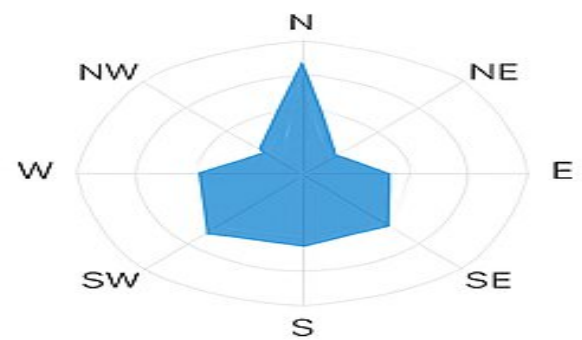
Monthly Rainfall



Weather by Region



Wind Direction



2. Information Visualization

Definition

- The graphical representation of **abstract data that does not have a natural physical form**.
- Focuses on large, complex, non-numerical datasets like text, networks, or categories.

Characteristics

- Deals with **abstract, categorical, relational, or textual data**.
- Data may be discrete rather than continuous.
- Goal is to **reveal patterns, relationships, and insights**.
- Techniques emphasize **interaction** (filtering, zooming, brushing, linking).

Examples

- Social network graphs (Facebook, LinkedIn connections).
- Financial dashboards (profit trends, stock movements).
- E-commerce analytics (customer segments, product performance).
- Text visualization (word clouds, topic clusters, sentiment analysis).

Example Scenario: Social Media Network Analysis

•Suppose we collect data about people and their connections on a platform like Facebook or LinkedIn.

•Raw data may look like:

- User ID, Name, Age, Country
- Friend Connections (who is connected to whom)

•In a visualization:

- Each person can be shown as a **node (circle)**.
- Friendships are shown as **lines (edges)** connecting nodes.
- The size of the circle can represent number of friends, and color can represent country or age group.

Why this is Useful?

•You can easily spot **influential people** (those with the most connections).

•You can identify **clusters or groups** (e.g., people from the same school or workplace).

•Businesses can use this to target **key influencers** for marketing campaigns.

Information Visualization = Taking **abstract datasets** → turning them into **graphs, charts, or networks** → so humans can **see patterns and make better decisions**.

Power BI Dashboard – Social Media Network Analysis

Example Dataset Fields

User Data

- User ID
- Name / Handle
- Age, Gender, Location

Network Data

- Followers Count
- Following Count
- Connections (User A → User B)
- Groups/Communities

Engagement Data

- Posts Shared
- Likes
- Comments
- Shares
- Engagement Rate

User ID	Name	Region	Gender	Age Group	Followers	Following	Posts Shared	Likes	Comments	Shares
1	User1	North America	Male	18-24	5230	300	120	4300	600	250
2	User2	Europe	Female	25-34	8920	4200	85	2500	320	180
3	User3	Asia	Male	35-44	3150	980	65	1500	200	90
4	User4	Other	Female	45-54	6720	2700	150	4100	750	320
5	User5	Europe	Male	25-34	4800	1100	40	1200	150	75

Step 1: Import Data

- Get social media dataset (CSV/Excel/SQL or API extract).
- Import into **Power BI Desktop** → *Get Data*.
- Ensure relationships between **Users Table** and **Posts Table** are created.

Step 2: Create Measures (DAX)

```
Total Followers = SUM(Users[Followers Count])
Total Engagement = SUM(Posts[Likes]) +
SUM(Posts[Comments]) + SUM(Posts[Shares])
Engagement Rate % = DIVIDE([Total Engagement],
SUM(Posts[Posts Shared]), 0)
Top Influencers = DISTINCTCOUNT(Users[User ID])
```

Step 3: Build Visuals

1.KPI Cards

1. Total Followers
2. Total Engagement
3. Engagement Rate %
4. Active Users

2.Network Graph (Custom Visual)

1. Nodes = Users
2. Links = Connections (Followers/Following)
3. Node size = Followers count
4. Node color = Region / Community

3.User Engagement Analysis

1. Bar Chart → Top 10 Influencers (by followers or engagement).
2. Line Chart → Engagement trend over time.

4.Content Analysis

1. Column Chart → Posts shared vs engagement.
2. Word Cloud → Most common hashtags or topics.

5.Demographics

1. Pie Chart → User distribution by Gender.
2. Map → Users by Country/Region.

Step 4: Dashboard Layout

- **Top Row** → KPI Cards (Followers, Engagement, Rate, Active Users).
- **Center** → Network Graph (main visualization).
- **Right Side** → Top Influencers Bar Chart + Engagement Trend Line Chart.
- **Bottom Row** → Map (Users by Location) + Word Cloud (Hashtags).
- **Left Side** → Slicers (Region, Gender, Age Group).

Step 5: Insights Gained

- Who are the **top influencers** in the network?
- How do **engagement trends** change over time?
- Which **regions/communities** are most active?
- What **content/hashtags** drives the most interaction?
- How are users **connected** and which clusters exist?

Social Media Network Analysis

Total Followers	Total Engagement	Engagement Rate %	Active Users
1.2M	2.4M	3.2%	85K

Region

- ☐ North America
- ☐ Europe
- ☐ Asia

Gender

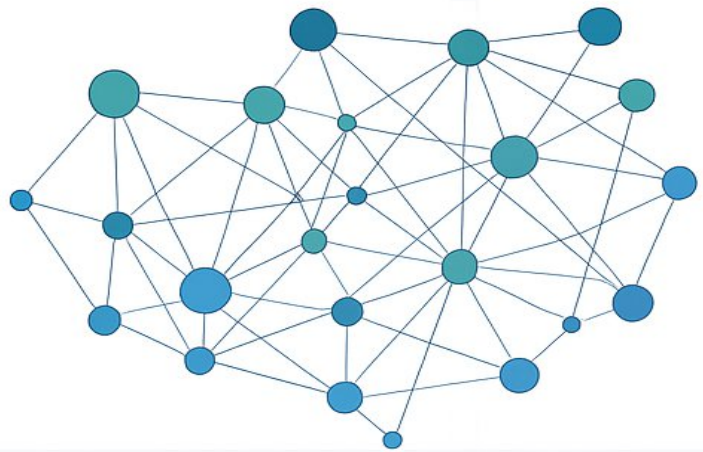
- ☐ Male
- ☐ Female

Age Group

- ☐ 18-24
- ☐ 35-34
- ☐ 52-34
- ☐ 45-34

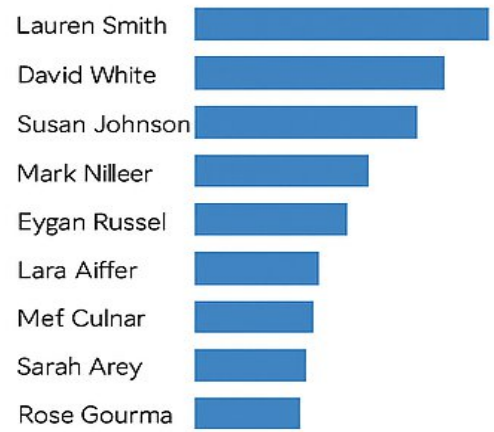
User Demographics





- Niewreed
- Eutope
- Asia
- Other

Top Influencers



Engagement Trend



Top Hashtags

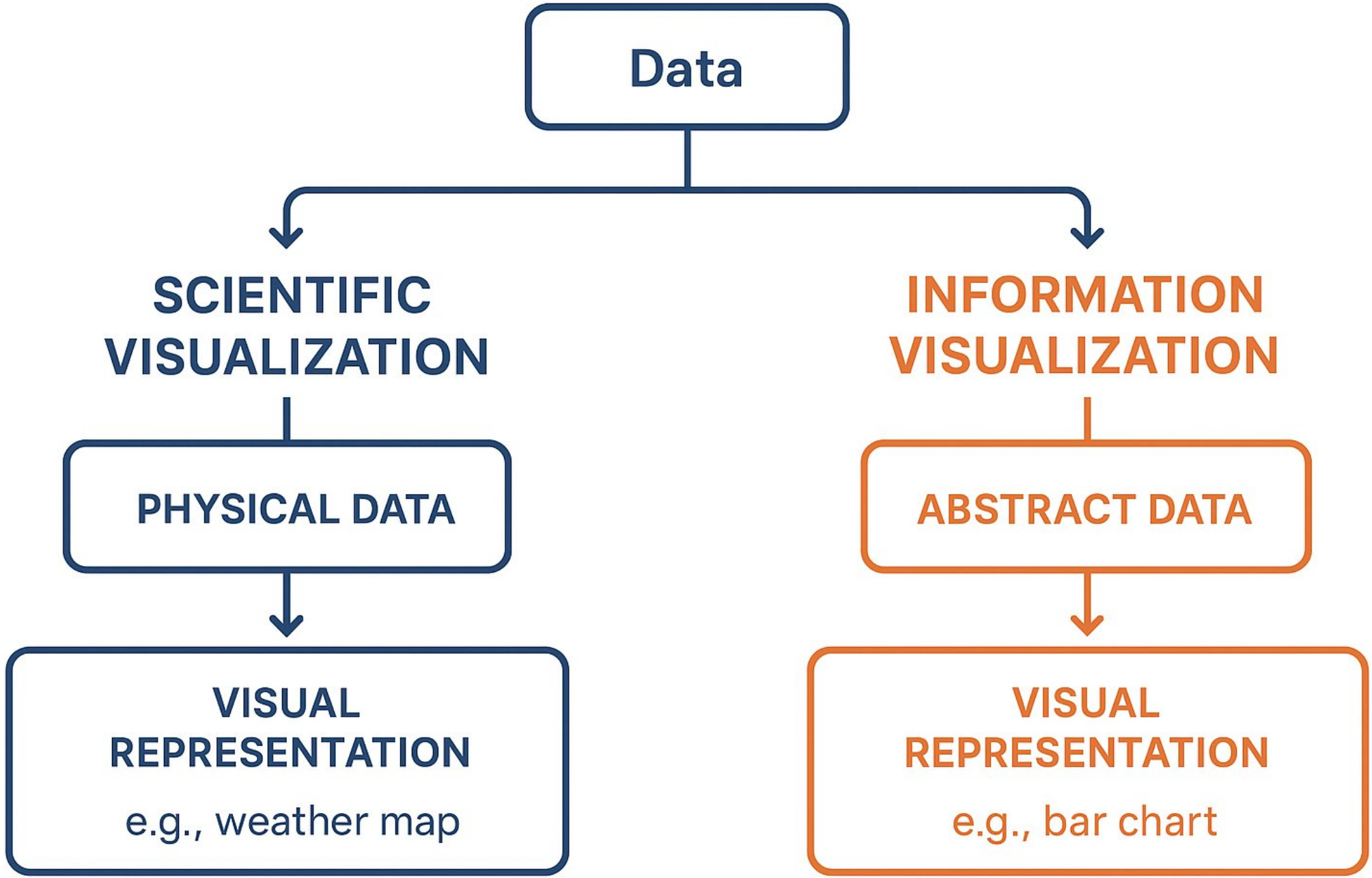


Users by Location



Comparison Table

Feature	Scientific Visualization	Information Visualization
Data Type	Physical, real-world, continuous data	Abstract, categorical, relational, discrete data
Source	Sensors, simulations, experiments	Databases, text, logs, networks
Goal	Understand physical phenomena (simulation, prediction)	Discover patterns, trends, and relationships in abstract data
Techniques	Volume rendering, isosurfaces, vector fields, medical imaging	Charts, graphs, treemaps, network diagrams, word clouds
Examples	Weather modeling, fluid dynamics, MRI scans	Social networks, business dashboards, text analysis


Data
SCIENTIFIC VISUALIZATION
PHYSICAL DATA
VISUAL REPRESENTATION
e.g., weather map
INFORMATION VISUALIZATION
ABSTRACT DATA
VISUAL REPRESENTATION
e.g., bar chart

Interaction techniques: filtering, zooming, brushing, linking,

Filtering

- **Definition**: Selecting a subset of data and hiding the rest to reduce clutter.
- **How it works**: Users apply filters (like year, region, category) to focus only on relevant information.
- **Example in Power BI/Tableau**: Using slicers to display sales data only for "2023" or "Technology" category.
- **Benefit**: Helps deal with **large datasets** by showing only what is important.

Example:

- **E-commerce Dashboard (Amazon, Flipkart, etc.)**
- A manager filters sales data to show only **“Electronics Category” in 2023**.
- All other data (clothing, furniture, etc.) is hidden.

Real-Time Use: Quickly focus on **specific product categories, time periods, or regions**.

Zooming

- **Definition**: Navigating between different levels of detail by enlarging or reducing the view.
- **Types**:
  - Geometric zoom (like Google Maps – zoom into cities or zoom out to the world).
  - Semantic zoom (changing the type of information shown when zooming, e.g., bar chart changes into table).
- **Example**: In a sales dashboard, zoom into "Asia" → then "India" → then "Bangalore" for more detail.
- **Benefit**: Supports **hierarchical data exploration**.

Example:

- **Google Maps / Weather Maps**
- Start from a **world map → zoom into Asia → India → Karnataka → Bangalore**.
- Each zoom level gives **more details** (global view → country stats → city weather).

Real-Time Use: Helps in **geographic or hierarchical data analysis**.

Brushing

- **Definition**: Highlighting data points in one visualization by selecting them directly.
- **How it works**: Click-drag or hover over a group of points → they get highlighted.
- **Example**: In a scatter plot, brushing over high-profit customers highlights them while dimming others.
- **Benefit**: Lets users **focus attention** without removing context.

Example:

- **Stock Market Analysis (NSE/BSE/Dow Jones Dashboards)**
- On a scatter plot of companies (Profit vs. Market Cap), the analyst brushes (highlights) the **top 10 companies by revenue**.
- Those companies get highlighted while others fade in the background.

Real-Time Use: Helps in **focusing on interesting patterns or outliers**.

Linking

•**Definition**: Connecting multiple visualizations so that interaction in one updates the others.

•**How it works**: Select a region in a map → related sales appear highlighted in bar and line charts.

•**Example**: Selecting "California" on a map updates the profit trend line and top customer bar chart automatically.

•**Benefit**: Enables **coordinated multiple views** of the same dataset.

Example:

•**Power BI Sales Dashboard**

•Selecting **“California” on a map** automatically updates the:

- Line Chart (sales trend in California)
- Bar Chart (top products sold in California)
- KPI Card (total California revenue).

Real-Time Use: Makes dashboards **dynamic and connected**.

- **Filtering** = Focus on subset (E-commerce filters)
- **Zooming** = Drill into details (Maps, Weather)
- **Brushing** = Highlight subset (Stock market scatterplot)
- **Linking** = Synchronize multiple visuals (Power BI dashboards)

Summary Table

Technique	What it does	Example	Benefit
Filtering	Shows only selected subset	Filter sales for 2023 only	Reduces clutter
Zooming	Moves between detail levels	Zoom into country → state → city	Hierarchical exploration
Brushing	Highlights selected items	Highlight top 10 customers in scatter plot	Focus attention
Linking	Connects multiple views	Select region → updates map & chart	Coordinated analysis

INTERACTION TECHNIQUES

Filtering

Show only selected subset



→ Filter sales for 2023 only

Zooming

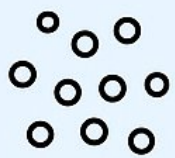
Navigate detail levels



→ Zoom into country → state → city

Brushing

Highlight selected items



→ Higghlight top 10 custo-

Linking

Connect multiple views



→ Select region → update map & chart

Misleading visualizations – how to avoid bias and distortion

Why Visualizations Mislead

Visualizations can mislead either **intentionally** (manipulating perceptions) or **unintentionally** (through poor design choices). Misleading visuals can distort reality and bias interpretation, making accurate decision-making harder.

Common Misleading Techniques

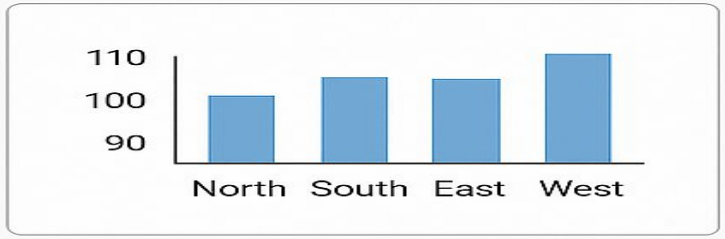
Technique	Description	Why It's Misleading
Truncated Axes	Y-axis (or X-axis) does not start at zero	Exaggerates small differences
Inconsistent Scaling	Axes use different scales or non-linear intervals	Misrepresents relative values
3D Charts	Use of perspective to display bar or pie charts	Distorts sizes and makes comparisons harder
Omitting Context	Only showing a part of the data (e.g., favorable time range)	Hides the bigger picture
Cherry-Picked Categories	Selecting specific data points or categories	Skews reader's understanding
Improper Aggregation	Over-summarizing or combining categories without explanation	Misleads about underlying distributions
Pie Charts with Too Many Slices	Too many categories shown at once	Difficult to read and compare sizes
Overuse of Color/Effects	Excessive colors, gradients, or shadows	Distracts from actual data

Guidelines to Avoid Bias and Distortion

Best Practice	Implementation Tip
Use full axes (start at zero)	Avoid truncating unless it improves clarity and is clearly indicated
Maintain consistent scales	Especially when comparing charts side-by-side
Limit 3D usage	Prefer flat, clean 2D charts for accurate comparisons
Show full data ranges	Provide appropriate temporal or categorical context
Label clearly and fully	Axes, units, sources — all must be clearly marked
Choose the right chart type	Match data characteristics to appropriate visual encodings
Be transparent	Mention any data cleaning, transformation, or filtering

Guidelines to Avoid Bias and Distortion in Data Visualization

Misleading Scales



Correct



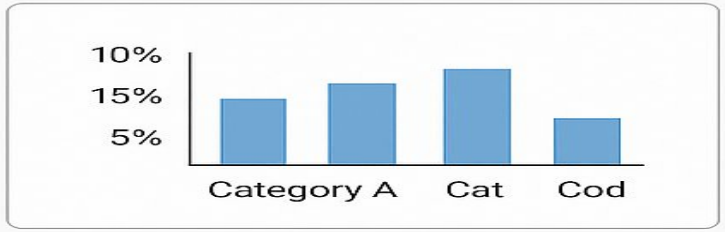
Cherry-picking Data

Profit = $5M

Correct

Revenue = $20M
Cost = $15M
Profit = $5M

Ambiguous Labels



Amounous



Emotional Colors/ Images



Emotional Colors/Images



example illustrating a **misleading visualization** and how to **correct** it

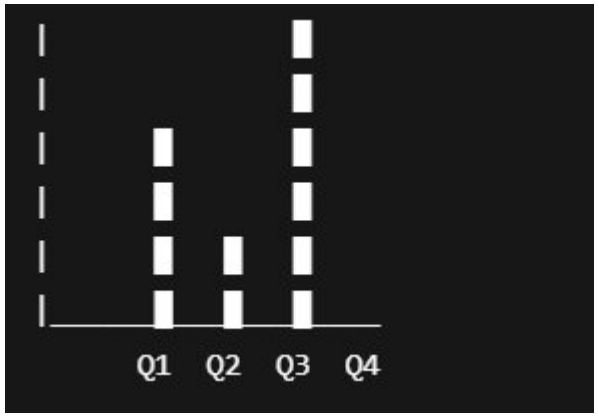
Truncated Y-Axis Misleading Bar Chart

Scenario:
A company wants to show its **quarterly sales growth** to investors.

Misleading Visualization (Truncated Y-Axis)

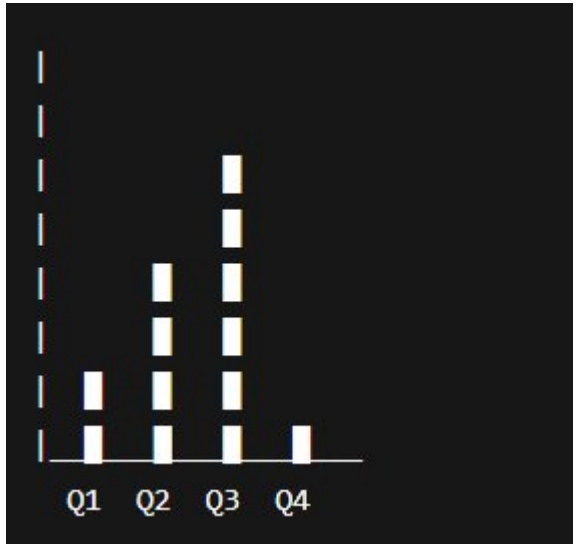
Quarter	Sales (in $ thousands)
Q1	100
Q2	105
Q3	107
Q4	110

Bar chart with Y-axis starting at 95k:



- The Y-axis **starts at 95k**, not zero.
- Bars look very different in height, exaggerating small changes.
- Gives the impression of rapid sales growth.

Correct Visualization (Y-Axis Starts at Zero)



- Y-axis starts at zero.
- Differences between quarters look modest and realistic.
- Gives a **true picture** of sales growth.

- The **truncated axis** misleads investors by making the growth seem much more dramatic.
- Starting Y-axis at zero preserves proportionality and avoids exaggeration.
- Adding axis labels and exact values further helps clarity.

Visualization of 1D (temporal),

- 1D data represents a **single variable changing over time** (temporal axis).
- Common in time-series analysis.

Examples:

1.Line Chart

1. Plots values (y-axis) over time (x-axis).
2. Example: Stock price trend, temperature variation over a week.

2.Timeline / Gantt Chart

1. Shows events or tasks along a temporal axis.
2. Example: Project management timelines.

3.Bar/Column Chart with Time Axis

1. Bars represent totals/counts at each time interval.
2. Example: Monthly sales, daily website visitors.

1D temporal visualizations are best suited for showing **trends, patterns, and seasonality** in data across time.

Line Chart (Temperature over Days)

- **What it shows:** Daily temperature readings over a week/month.
- **Example:**
  - X-axis = Days (Mon–Sun)
  - Y-axis = Temperature (°C)
  - Line goes up/down showing warming/cooling trends.
- **Use case:** Weather forecasting.

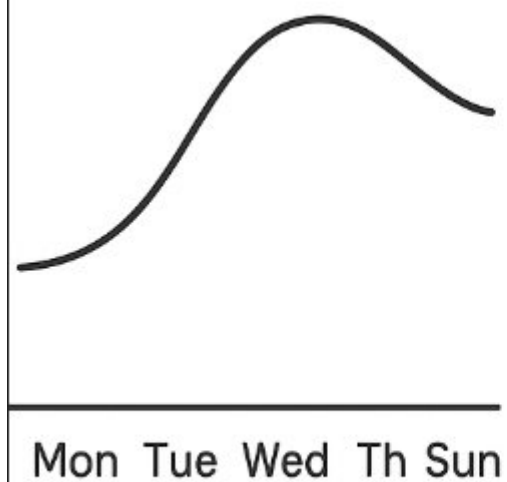
Stock Price Chart

- **What it shows:** Price fluctuations of a stock over time.
- **Example:**
  - X-axis = Dates (Jan–Dec 2024)
  - Y-axis = Stock Price (USD)
  - Line shows rises and falls (with possible moving averages).
- **Use case:** Finance/investment tracking.

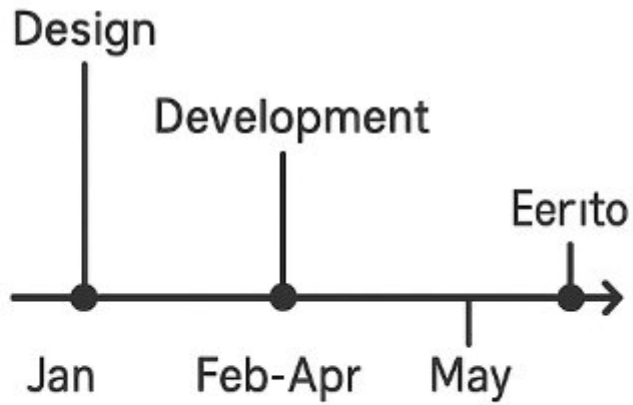
Timeline of Events

- **What it shows:** Sequence of events along a time axis.
- **Example:**
  - Project timeline → "Design (Jan), Development (Feb–Apr), Testing (May), Launch (June)".
  - Events placed on a straight line with labels.
- **Use case:** Project management, history timelines.

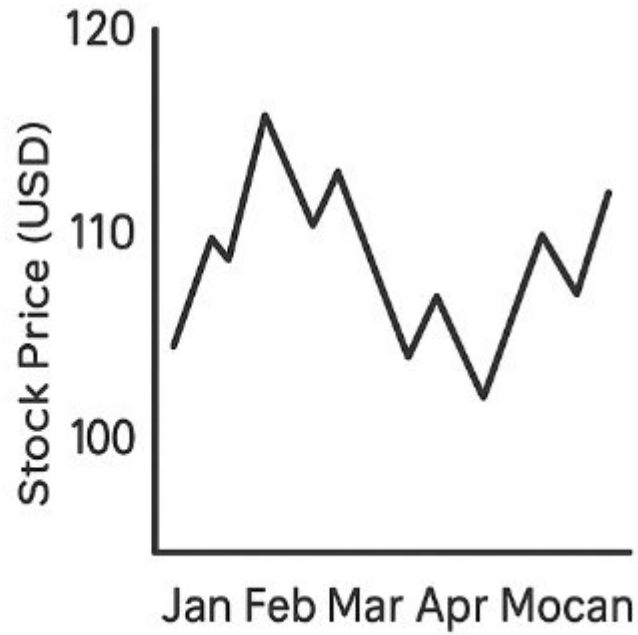
Temperature over Days



Timeline of Events



Stock Price Chart



2D (maps, scatterplots),

In **data visualization**, **2D visualizations** are the most commonly used because they map information on a two-dimensional plane (X-axis and Y-axis). They are powerful for showing relationships, distributions, and geographic patterns.

. Maps (Geographical Visualization)

•**What it is:** Visual representation of data over a geographical area (countries, states, regions, cities).

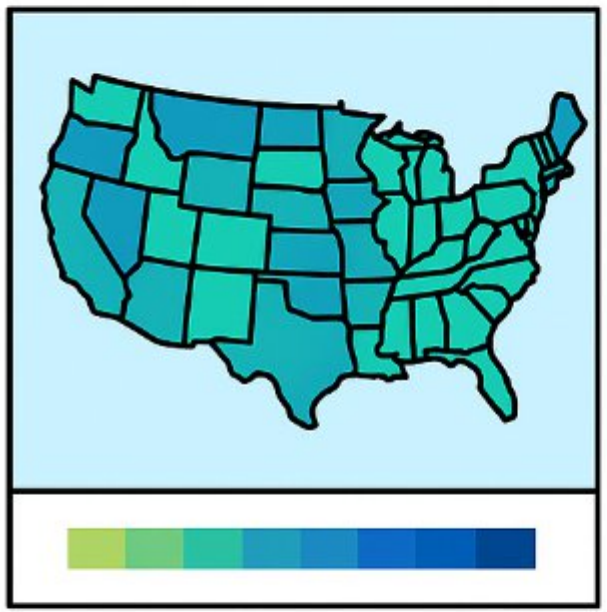
•**Purpose:** To show **spatial patterns, density, and comparisons** across locations.

•**Examples:**

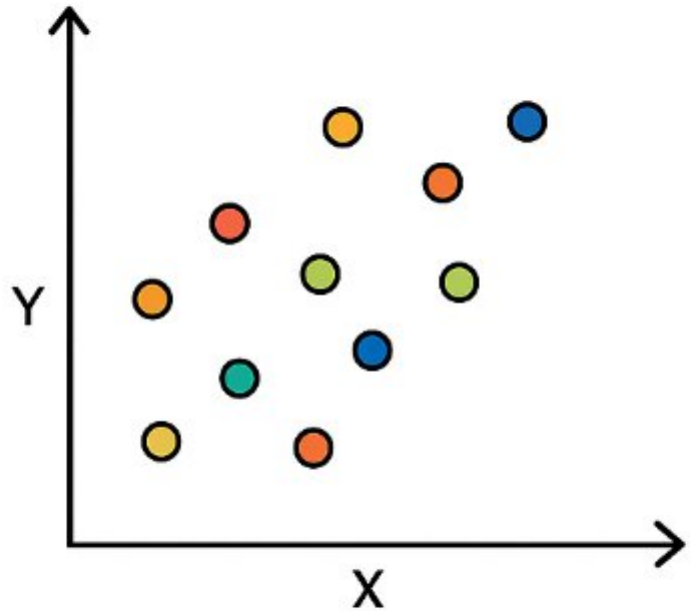
- Choropleth map (color-coded by population density, literacy rate, sales, etc.)
- Heat map (showing intensity of events like crime, disease spread, traffic congestion)
- Symbol map (circles of different sizes on a map for sales volume, earthquake magnitude, etc.)

📌 *Use case:* Showing **COVID-19 cases per country**, **sales by region**, or **rainfall distribution** on a map.

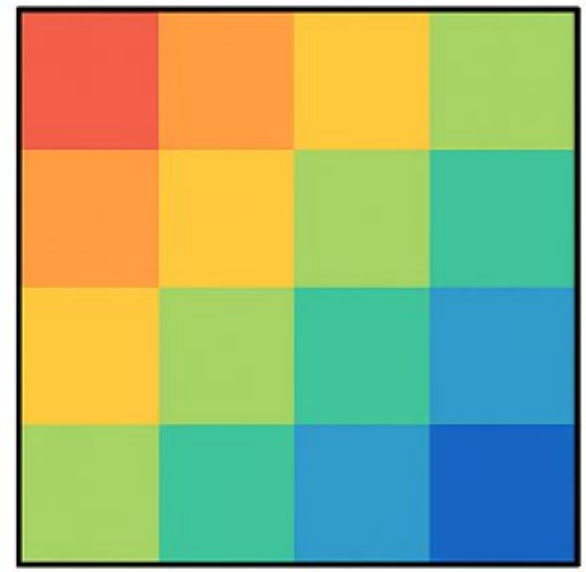
2D (maps, scatterplots)



Map



Scatterplot



Heatmap