Unit - II

Creating visual representations: visualization reference model, visual mapping, visual analytics, Design of visualization applications.

1. Visualization Reference Model

The **Visualization Reference Model**, often referred to as the **Visualization Pipeline**, is a conceptual framework that outlines the process of transforming raw data into meaningful visual representations. This model is widely used in data visualization to enhance understanding, exploration, and presentation of data.

The **Visualization Reference Model** describes the flow from **raw data** to **interpretable visual representation**.

Stages:

1. **Raw Data** → Collected from databases, files, sensors.
2. **Data Transformations** → Cleaning, filtering, aggregation.
3. **Visual Mapping** → Mapping data attributes to visual properties (e.g., x-axis = time, color = category).
4. **View Transformation** → Scaling, zooming, filtering within the visual.
5. **Visual Presentation** → Final chart, graph, or interactive display.

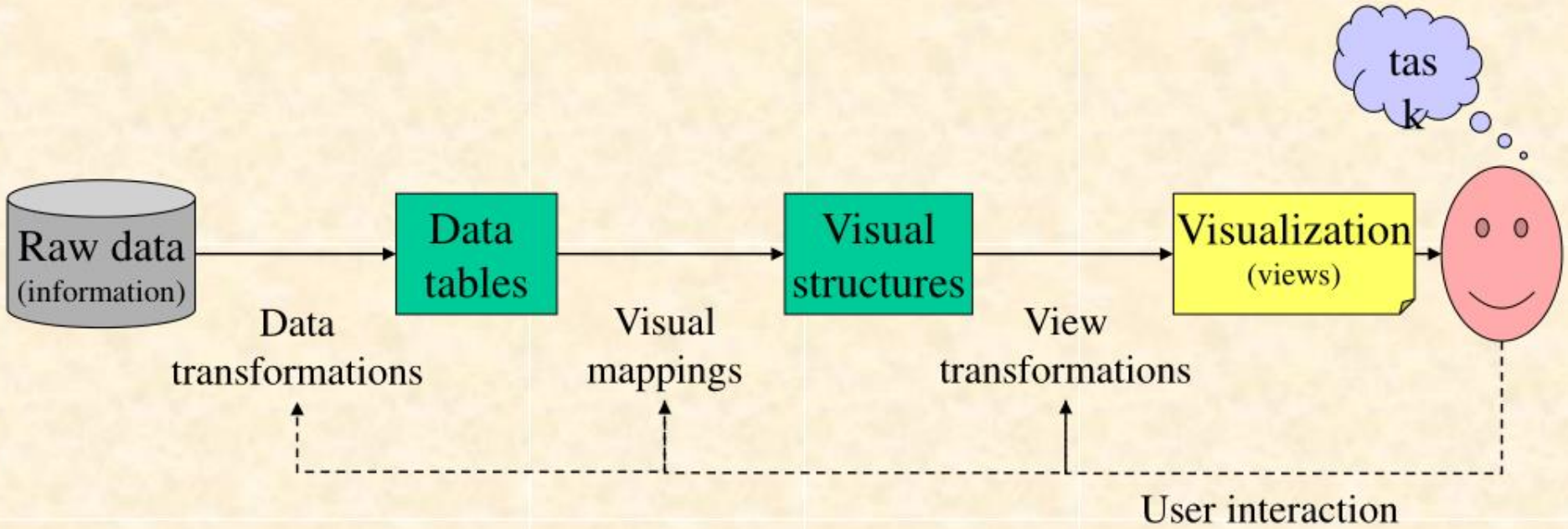
The pipeline consists of several key stages:
1.Data Preparation: This initial step involves preparing the raw data for visualization. Techniques such as interpolation or data analysis are applied to structure the data in a way that aligns with the visualization task.

2.Filtering: In this stage, users interactively select specific subsets of data to visualize. This step is particularly important in exploratory visualization, where users search for unknown patterns or features in the data.

3.Visualization Mapping: The prepared data is transformed from its original space (data space) into a visualization space. This involves creating abstract representations, such as 2D or 3D geometric shapes, that effectively convey the data's meaning.

4.Rendering: The visualization space is converted into an image or animation that can be displayed to the user. This step uses computer graphics techniques like shading and surface detection to produce the final visual output.

Visualization Pipeline



1. Raw Data (Information)

•**What it is:** The initial collected data in its original form.

•**Real-time example:**

In an e-commerce company like Amazon, raw data could be every transaction — order ID, product name, quantity, price, date, customer location.

•**Why it matters:** Without cleaning or structuring, this data isn't directly usable for analysis.

2. Data Tables (After Data Transformations)

•**What it is:** Data that has been processed, cleaned, and organized into a structured format like tables.

•**Real-time example:**

From the raw transactions, you might prepare a table showing **monthly sales by product category**, removing duplicates and correcting missing values.

•**Transformation:**

- Removing invalid entries (e.g., negative prices)
- Aggregating data (e.g., total sales per month)

3. Visual Structures (After Visual Mapping)

•**What it is:** Mapping data attributes to visual elements like bars, colors, shapes, or positions.

•**Real-time example:**

- Product categories → Different colors
- Monthly sales amount → Height of bars in a bar chart

•**Why it matters:** This step defines *how* the data will be visually represented so that it's easy to understand.

. Visualization (Views) — After View Transformation

•**What it is:** The final view presented to the user, possibly with interaction features like filters, zooming, and drill-downs.

•**Real-time example:**

- An interactive dashboard where managers can filter by year or click on a product category to see details.

•**View Transformations:** Adjusting scale, reordering items, changing chart type, zooming in/out.

5. User Interaction & Feedback Loop

•**What it is:** Users exploring and interacting with the visualization, leading to new questions and refined analysis.

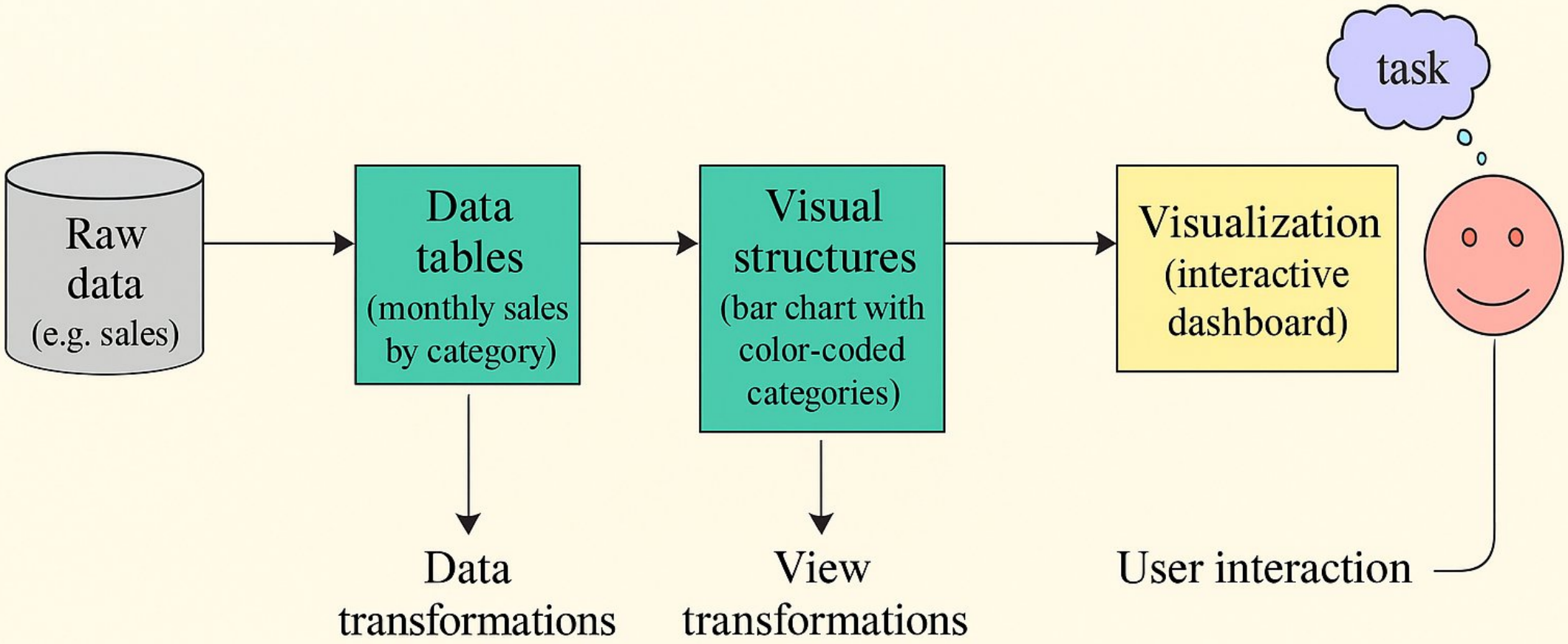
•**Real-time example:**

- A sales manager clicks on "Electronics" in the bar chart to see sales broken down by individual products.

•**Why it matters:** The feedback loop allows deeper insights and iterative exploration of data.

Stage	Example in E-Commerce	Process
Raw Data	Order details from all customers	Collected from database
Data Tables	Monthly sales by category	Cleaning & aggregation
Visual Structures	Mapping sales → bar height, category → color	Visual mapping
Visualization	Interactive dashboard	Filtering, zooming
User Interaction	Click category to drill down	Continuous analysis

Visualization Pipeline



Visualization Reference Model to a **Power BI example**

1. Raw Data (Information)
Theory: This is the unprocessed data collected from a source.
Power BI Example:
•Importing an Excel file of **sales transactions** with columns like *Date*, *Region*, *Product*, *Quantity Sold*, *Revenue*.
•Source could also be SQL, CSV, or an API.

Raw Data (CSV extract):

Date	Prod_ID	Region	Qty_Sold	Rev($)	Cost($)
2025-01-05	P01	East	50	5000	3500
2025-01-05	P02	North	20	2000	1500
2025-01-06	P01	West	30	3000	2100

2. Data Tables (After Data Transformation)

Theory: Cleaned, filtered, and structured data ready for analysis.

Power BI Example:

•Using **Power Query Editor** to:

- •Remove duplicates
- •Change data types
- •Merge tables (e.g., Sales with Product details)
- •Add calculated columns like Profit = Revenue – Cost

Transformations done:

•Joined **Product ID** with product details to get names.

•Calculated new column: **Profit = Revenue - Cost**

•Removed unused columns like Prod_ID, Cost($).

•Ensured correct data types for date, text, and numbers.

Data Table after transformation in Power Query:

Date	Product Name	Region	Quantity Sold	Revenue ($)	Profit ($)
2025-01-05	Laptop	East	50	5000	1500
2025-01-05	Mouse	North	20	2000	500
2025-01-06	Laptop	West	30	3000	900

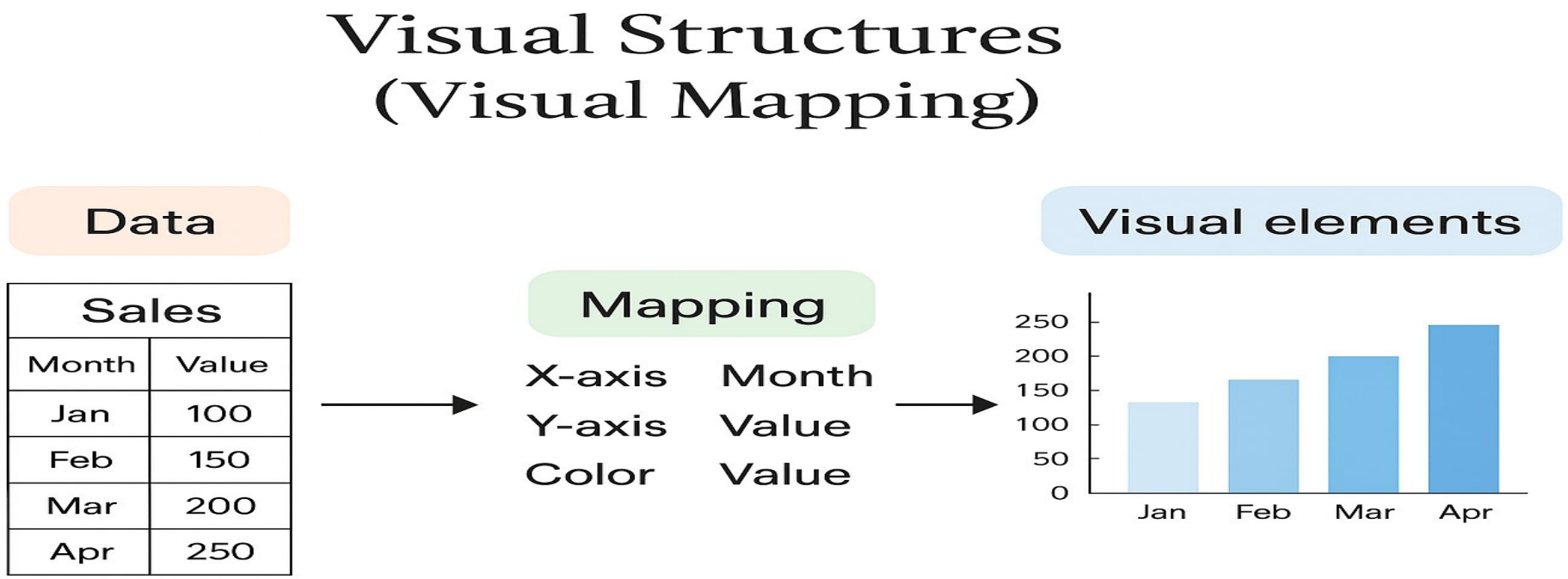
3. Visual Structures (Visual Mapping)

Theory: Mapping the data to graphical elements (bars, lines, colors, size, position).

Power BI Example:

•Creating a **bar chart** where:

- X-axis = Months
- Y-axis = Total Revenue
- Color = Region



4. Visualization (Views)
Theory: Final rendered visuals shown to the user for interpretation.
Power BI Example:
•Dashboard containing:

- **Revenue by Region (Map Visual)**
- **Monthly Sales Trend (Line Chart)**
- **Top 5 Products (Bar Chart)**

•Slicers to filter by year, region, or product category.

5. User Interaction
Theory: Users manipulate views to get insights.
Power BI Example:
•A manager clicks on *North Region* in the map visual, and all other charts update (cross-filtering) to show only *North Region* data.

Power BI Dashboard Example Setup

1. Data Preparation

Ensure your dataset has at least:

- **Region** (e.g., North, South, East, West)
- **Date** (monthly sales data)
- **Product Name** and **Category**
- **Revenue** or **Sales Amount**

Real-Time Example Flow in Power BI:

1.Import *Retail Sales Data (CSV)* → **Raw Data**
2.Clean in Power Query → **Data Tables**
3.Map "Region" to color, "Revenue" to height of bars → **Visual Mapping**
4.Display as bar chart + map → **Visualization**
5.Add slicers for year & category → **User Interaction**

A.Dashboard Components

A. Revenue by Region (Map Visual)

- **Visual Type**: Map
- **Location**: Region
- **Size or Color**: Revenue
- This shows where revenue is concentrated geographically.

B. Monthly Sales Trend (Line Chart)

- **Visual Type**: Line chart
- **X-axis**: Month/Date
- **Y-axis**: Revenue
- **Legend**: Region (optional for regional trends)
- This reveals seasonality or sales patterns over tim

c. Top 5 Products (Bar Chart)

- **Visual Type**: Clustered Bar Chart
- **X-axis**: Revenue
- **Y-axis**: Product Name
- Apply **Top N Filter** to show only Top 5 products by revenue.

D. Slicers

- Add slicers for:
  - **Year**
  - **Region**
  - **Product Category**
- This enables interactive filtering for different perspectives.

ar

:23

Region

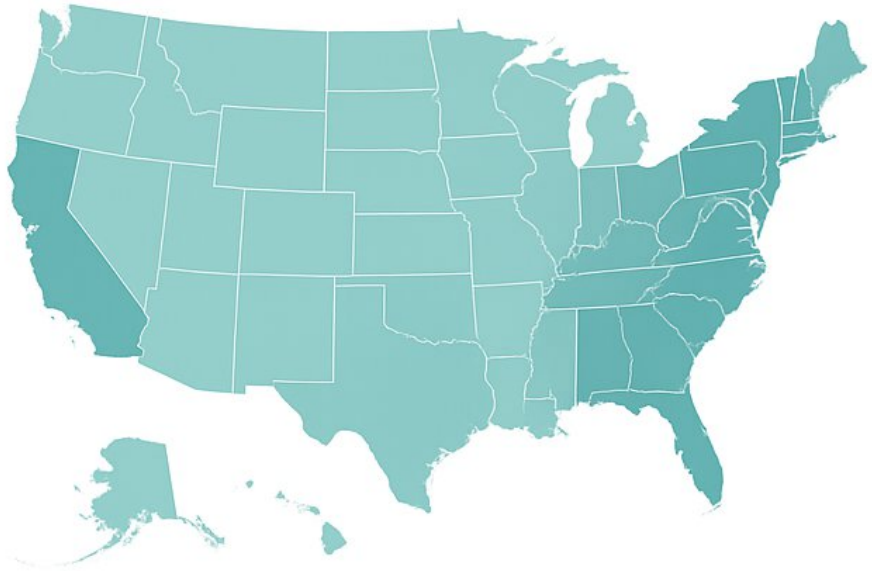
East

Product Category

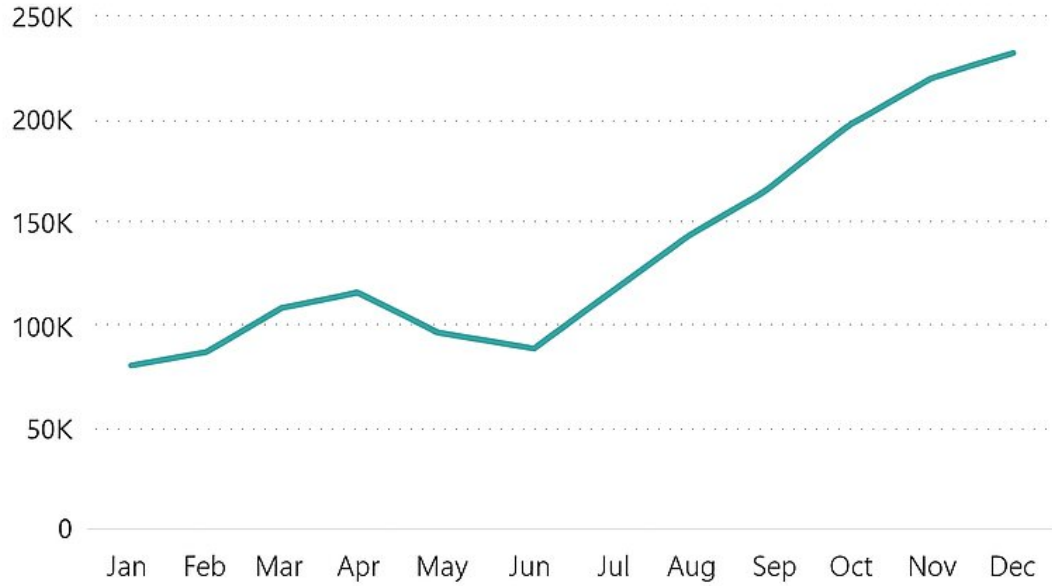
East

West

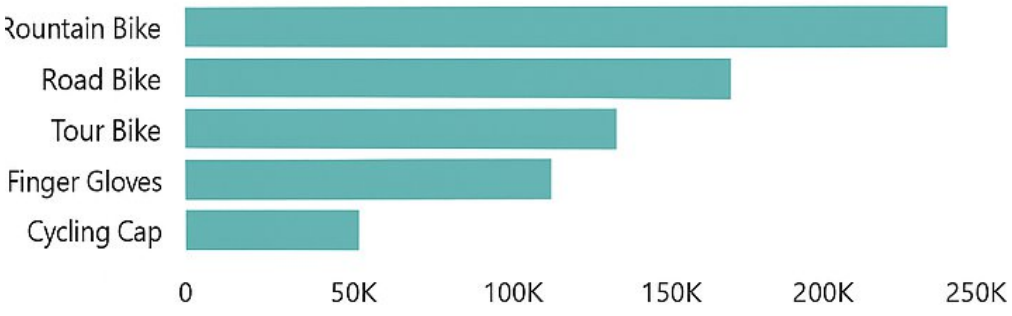
evenue by Region



Monthly Sales Trend



op | 5 Products



Year

2023

Region

East

North

South

Product Category

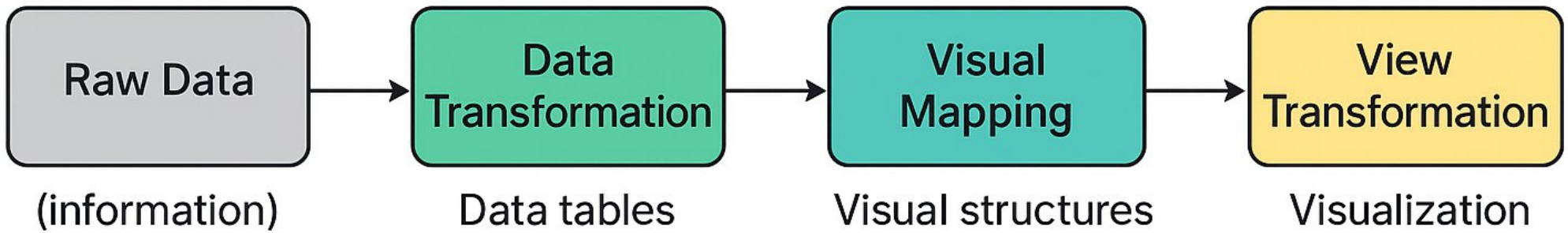
Accessories

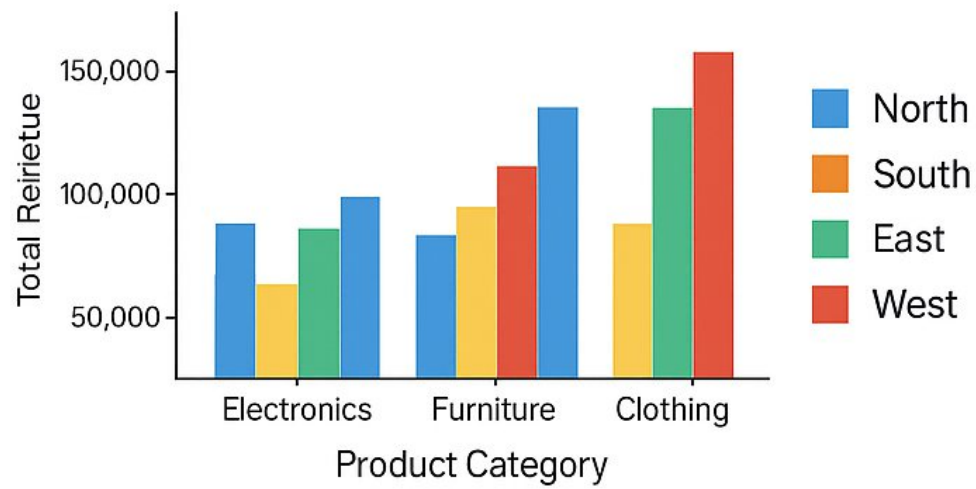
Bikes

Clothing

Components

Visualization Pipeline





Visual Mapping

Visual mapping is the process of **transforming data values into visual elements (marks)** using **visual variables** such as position, size, color, shape, orientation, and texture.
It is a key step in the **Visualization Pipeline** — converting processed data into a visual representation that people can interpret.

This technique is widely used in various fields, including education, business, and design, to facilitate brainstorming, planning, and problem-solving.

Purpose

- Makes complex data easier to understand.
- Highlights patterns, trends, and outliers.
- Allows comparisons between data points visually.

The Three Elements of Visual Mapping for Information Visualization

- The Spatial Substrate
- The Graphical Elements
- The Graphical Properties

The spatial substrate is the space in which we're going to create our visualization. The vast majority of information visualization takes place in 2-dimension space – along x and y axes. However, it is possible to create 3-dimensional and even hyper-dimensional (more than 3-dimensions) representations too.

It's important to consider which axes you will use and what kind of data will map to each individual axis.

Data may be:

- **Quantitative**. E.g. it has a numerical measurement of some quality of the data set.
- **Ordinal** e.g. is not associated with a number but still offers some form of order – for example months in a calendar year.
- **Nominal** e.g. the data is a collection of non-numerical and non-ordered data (the names of the members of parliament for example)

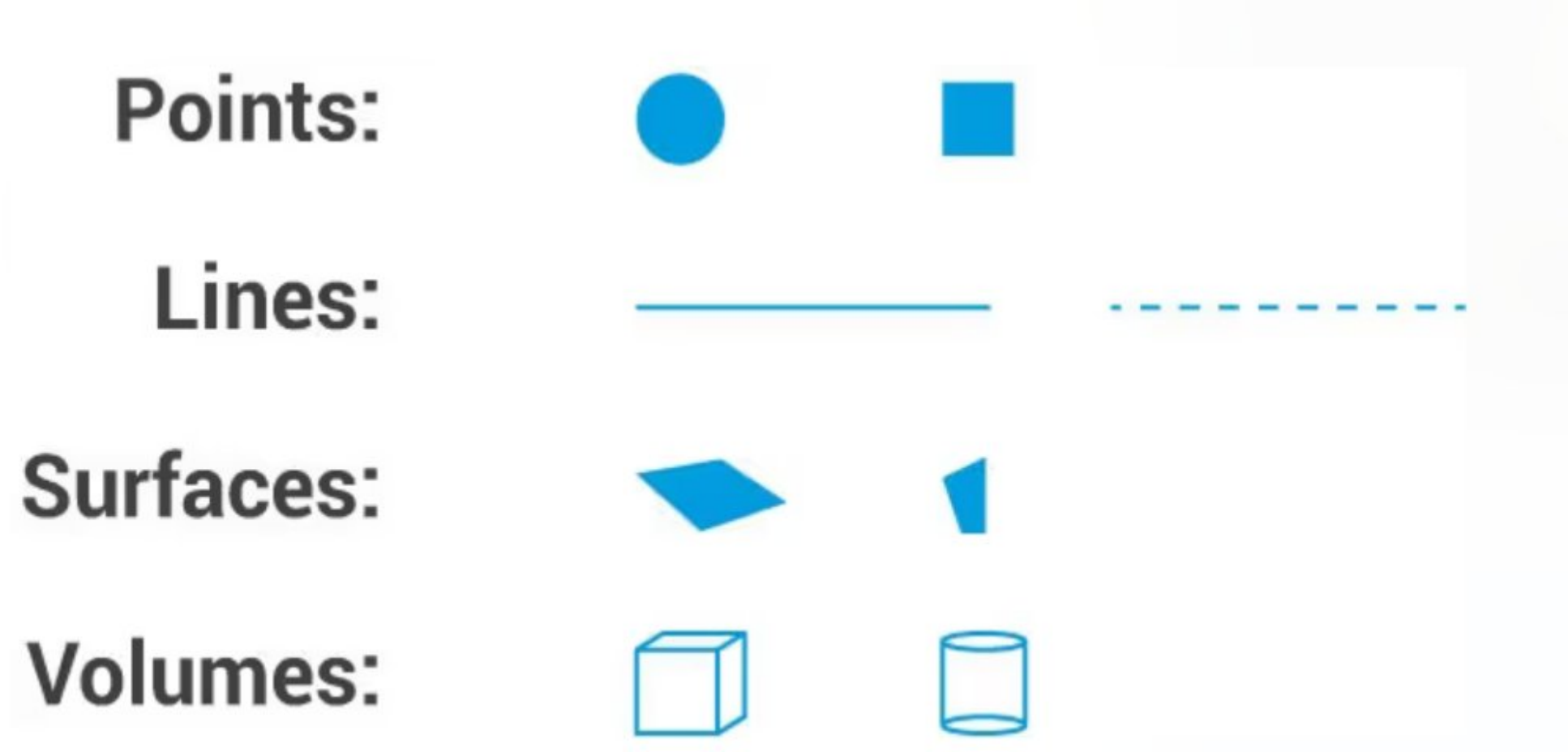
Data Types (Ward et al.)

1. **Categorical (Nominal)**
    1. Discrete, unordered labels (e.g., country, gender, product type).
    2. No inherent ranking.
2. **Ordinal**
    1. Ordered categories (e.g., ratings: poor → fair → good → excellent; ranks).
    2. Relative position matters, but not precise numeric difference.
3. **Quantitative**
    1. Numeric, continuous or discrete (e.g., height, revenue, age).
    2. Both magnitude and interval/ratio differences matter.

The Graphical Elements

Graphical elements are the visual elements that will appear in the spatial substrate. There are four kinds of visual elements:

- Points
- Lines
- Surfaces
- Volumes



3.Graphical properties are properties that can be applied to the graphical elements which make them more (or indeed less) noticeable to the eye and/or valuable to the user of the representation.

There are many possible graphical properties that can be applied to graphical elements but the most commonly used are:



Main Visual Mapping Techniques

Visual Variable	How it's Used	Example in Power BI
Position	Map data points to specific x/y coordinates.	Scatter plot (Sales vs. Profit)
Size	Larger marks represent bigger values.	Bubble chart showing revenue size
Color (Hue & Intensity)	Different colors or shades represent categories or magnitude.	Heatmap (dark color = high value)
Shape	Different shapes represent different categories.	Pie chart with slices
Orientation	Angle or direction shows trend or category.	Arrow indicators in KPI cards
Texture/Pattern	Different fill patterns indicate groups.	Map with textured regions

Visual Mapping Process

1. **Input:** Processed data table.
2. **Encoding:** Choose suitable visual variables.
3. **Rendering:** Apply these encodings to create charts, maps, dashboards.
4. **Interaction:** Add slicers/filters for user control.

visual channels (a.k.a. retinal variables or encoding variables) that can map data types to visual properties.
Examples: **position, length, angle, area, shape, color (hue & intensity), orientation, texture, motion**.

Example

- **Categorical data**: Countries → encoded by **color hue** or **symbol shape** on a map.
- **Ordinal data**: Customer satisfaction (low → high) → encoded by **ordered color gradient** or **size ranking**.
- **Quantitative data**: Sales revenue → encoded by **bar length** or **position on axis**.

Mapping Data Types to Visual Channels

Data Type	Suitable Visual Channels	Notes
Categorical	- **Position (x/y location)** - **Shape** - **Hue (color category)** - **Spatial region / enclosure**	Must emphasize *distinctness*; hue and shape help distinguish groups, but no ordering implied.
Ordinal	- **Position (ordered axis)** - **Saturation / lightness (ordered color)** - **Texture density** - **Size (ranked symbol size)**	Must preserve *order* without implying precise magnitude (e.g., light → dark shades for "low → high").
Quantitative	- **Position (precise, axis-based)** - **Length** - **Angle / slope** - **Area** - **Color intensity (gradient)**	Encodings must support accurate comparisons of magnitude and proportion. Position and length are most accurate.

Visual Mapping & Channel Accuracy Cheat-Sheet

1. Visual Mapping of Data Types to Visual Channels (Ward et al.)

Data Type	Best Visual Channels
Categorical	Color Hue Shape Position (separate regions) Enclosure
Ordinal	Position (ordered axis) Saturation / Lightness Size (ranked) Texture density
Quantitative	Position (precise axis) Length Angle / Slope Area Color Intensity (gradient)

2. Ranking of Visual Channels by Accuracy (Cleveland & McGill, cited in Ward et al.)

Rank	Visual Channel (Quantitative Accuracy)
1	Position (common scale)
2	Position (nonaligned scale)
3	Length
4	Angle / Slope
5	Area
6	Volume / Curvature
7	Shading / Color Saturation

Visual analytics

The science of combining automated data analysis techniques with interactive visualizations, so humans and machines work together to make sense of complex data.

Ward et al. (and also Thomas & Cook, 2005) describe it as:
“The synergistic use of data analysis and visualization for effective analytical reasoning.”

- **Data Analysis (Machine strength)**
  - Handles *large-scale computation*, statistics, machine learning, data mining.
  - Can extract patterns, clusters, correlations, anomalies.
- **Visualization (Human strength)**
  - Provides *visual representation* of the results.
  - Lets humans detect patterns, interpret results, ask “why,” and guide further analysis.

Together → **interactive exploration loop**:

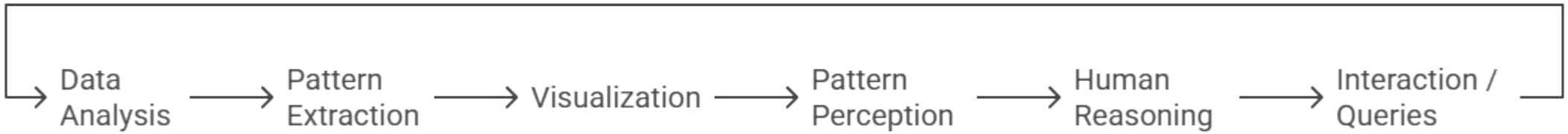
1. System computes → finds structure in data.
2. Visualization shows → human perceives patterns.
3. Human interprets → asks questions, refines filters/queries.
4. System updates analysis → new results visualized.

Workflow of Visual Analytics

1. **Data preprocessing** (cleaning, integration).
2. **Automatic analysis** (ML, statistics, clustering, anomaly detection).
3. **Visualization** (interactive dashboards, scatterplots, networks, maps).
4. **Human reasoning** (domain expertise, decision making).

This is often shown as a **pipeline with feedback loops** (human ↔ computer).

Data Analysis and Visualization Process



Example Scenarios

- **Cybersecurity**: anomaly detection algorithms highlight suspicious traffic → analyst explores visually.
- **Healthcare**: ML model predicts risk groups → dashboard lets doctors compare patient cohorts.
- **Business intelligence**: Sales clustering + interactive dashboards → managers spot trends quickly.

Benefits

- Handles **big, complex, uncertain data**.
- Combines **precision of algorithms** with **intuition of humans**.
- Supports **exploratory analysis** and **decision-making**.

Principles of Effective Visualization Design

1.Match Data Type to Visual Channel

- **Categorical (nominal):** encode with **hue, shape, position (separated regions)**.
- **Ordinal:** encode with **ordered hue/saturation, size ranking, position along ordered axis**.
- **Quantitative:** encode with **position, length, angle, area, color intensity**.

✓ Ensures proper mapping between **data semantics** and **visual encoding**.

2. Use the Most Accurate Visual Channels

•Cleveland & McGill ranking:

- Position on common scale → most accurate
- Length → next
- Angle / Slope → moderate
- Area, Color Intensity → less accurate
- Volume, Curvature → least accurate

  ✔ Prefer **position & length** for quantitative precision.

3. Reduce Cognitive Load

•Avoid clutter → don't overload with too many marks/colors.

•Use **Gestalt principles** (proximity, similarity, enclosure, continuity) to group related elements.

•Use **pre-attentive attributes** (color, size, shape) for fast recognition.

4. Maximize Data-Ink Ratio (Tufte)

•Remove unnecessary decoration (chartjunk).

•Focus ink on data, not gridlines, backgrounds, or 3D effects.

✓ Minimalist design improves clarity.

5. Support Comparisons

•Align items along a **common axis** when possible (bar charts > pie charts).

•Use consistent scales across related charts.

•Provide **context** (reference lines, baselines).

6. Use Color Effectively

•Use **hue** for categorical differences.

•Use **saturation/lightness gradients** for ordinal or quantitative values.

•Be mindful of **colorblind-friendly palettes**.

✓ Avoid misleading color schemes (e.g., rainbow).

7. Maintain Consistency

•Same variable → same visual encoding across charts.

•Consistent scales, legends, and interaction behaviours.

✓ Helps users build a mental model.

8. Enable Interaction (in VA)

•Filtering, zooming, brushing, linking across views.

•Supports exploration and hypothesis testing.

✓ Especially important in **Visual Analytics**.

9. Provide Annotations & Context

•Titles, axis labels, legends, units, tooltips.

•Show **uncertainty/error bars** where relevant.

✓ Guides interpretation, avoids misreading.

10. Know Your Audience & Task

•Visualization must be tailored to:

- **Audience:** Expert analyst vs. general public.
- **Task:** Exploration vs. presentation.

  ✓ Same data may need very different visualizations.

Effective visualization design = Right mapping + Accurate channels + Minimal clutter + Context + Audience focus.

Copy code


Principles of Effective Visualization Design
Data – Channel Mapping
Perception & Cognition
Design & Clarity
• Match data type → channel
• Position > Length > Area
• Gestalt principles
• Pre-attentive features
• Reduce clutter
• Maximize data-ink ratio
• Consistent encoding
• Effective use of color
Context & Audience
• Titles, labels, legends
• Comparisons (axes, baselines)
• Interactivity (zoom, filter, brush)
• Tailor to task & audience

Case Study 1: **Hospital Operations Dashboard**

A large hospital wants to monitor bed occupancy, ER wait times, and staff allocation.

Design Choices:

- **KPIs at top** (bed occupancy %, average ER wait time) → clear, large cards.
- **Color coding** → red for critical overload, green for safe zone.
- **Drill-down** → click ward → see patient-level details.
- **Visual Mapping:**
    - Quantitative → bar/line charts (trends of admissions).
    - Categorical → color-coded department filters.

Impact:

Faster decisions, reduced ER congestion, optimized staff deployment.

Hospital Operations Dashboard – Design

1. Layout Structure

•Top Section (Header KPIs)

- *Large KPI Cards* (Bed Occupancy %, ER Wait Time, Staff on Duty, ICU Availability)
- **Visual Style:** Bold numbers, color-coded (Green = Normal, Orange = Warning, Red = Critical).

•Left Panel (Filters/Controls)

- Department dropdown (General, Surgery, ICU, Pediatrics, etc.)
- Date/Time range picker
- Staff role filter (Doctors, Nurses, Support)

•Main Body (Visualizations)

- **Bed Occupancy Trend (Line Chart):** Admissions vs. Discharges over time.
- **ER Wait Times (Bar Chart):** Average wait time per department.
- **Ward/Department Status Map (Heatmap or Grid):** Color-coded beds (Green = Free, Red = Occupied).
- **Staff Allocation (Stacked Bar Chart):** Number of doctors/nurses per ward.

•Right Panel (Drill-Down & Alerts)

- Selected Ward → Patient List (Name, Age, Status, Admission Time).
- **Alerts Panel:**
  - "ICU occupancy at 95%" → Red warning.
  - "ER wait > 45 min" → Orange alert.

2. Visual Mapping Principles

- **Quantitative →** Line/Bar charts (trends & averages).
- **Categorical →** Color-coded filters (departments).
- **Status/Alerts →** Color + Icons (✓ normal, ⚠ warning, ✕ critical).
- **Hierarchy →** KPIs at top, details at bottom (overview → drill-down).

3. Design Features

- **Minimal clutter** → clean background, grid alignment.
- **Consistent color palette** → medical theme (Blue = info, Green = safe, Red = critical).
- **Interactivity** → drill-down, hover tooltips, time-range zoom.
- **Responsiveness** → scalable for desktop & tablet (for doctors on the go).

4. Expected Impact

- ✓ Instant overview of hospital capacity.
- ✓ Faster resource reallocation.
- ✓ Data-driven decisions in critical scenarios.
- ✓ Improved patient flow & reduced ER congestion.

Hospital Operations Dashboard

BED OCCUPANCY %

72%

ER WAIT TIME

35 min

STAFF ON DUTY

154

ICU AVAILABILITY

80%

FILTERS

Department

General

Date

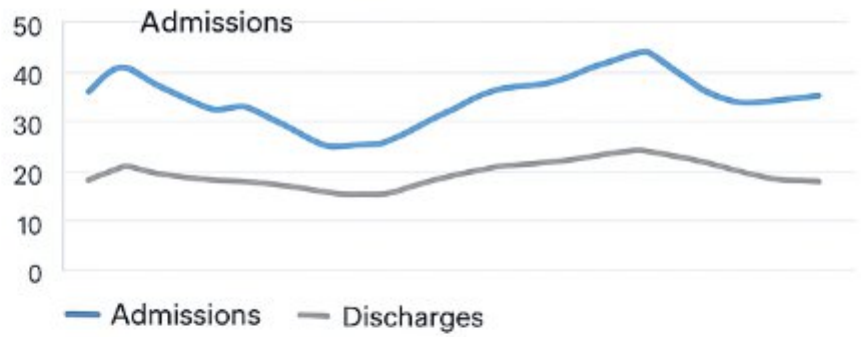
Staff

Doctors

Nurses

Support

BED OCCUPANCY TREND



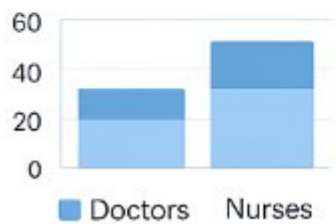
SELECTED WARD

Name	Age	Status	Alms
John Smith	34	In Surgery	
Emily Brown	-45	Under Observ	
Michael Davis	6	In Surgery	
Sarah Johnson		Pending	

WARD/DEPARTMENT STATUS



STAFF ALLOCATION



ALERTS

ICU occupancy at 95%

ER wait > 45 min

Case Study 2: E-Commerce Sales Dashboard

Context:

An online retailer wants insights into product performance and customer behavior.

Design Choices:

- **Sales by Region** → map visualization (geographic distribution).
- **Top Products** → horizontal bar chart (ranked).
- **Trends** → line chart (sales over time).
- **Interactivity** → slicers for category, time, location.
- **Design Principles:**
  - Reduce clutter (no unnecessary 3D).
  - Consistent colors per category (electronics = blue, clothing = green, etc.).

Impact:

Increased ability to identify best-selling products, seasonal demand patterns, and underperforming regions.

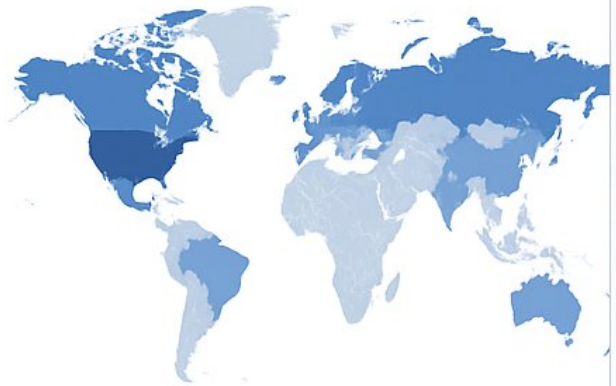
E-Commerce Sales Dashboard

Total Sales
$1.2M

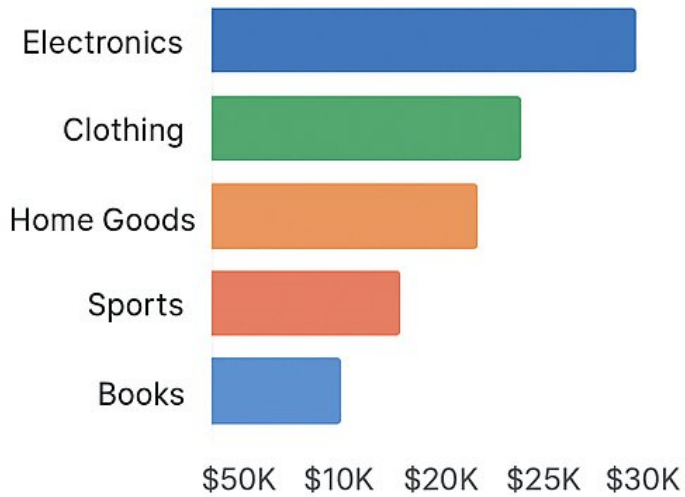
Avg. Order Value
$76

New Customers
1,580

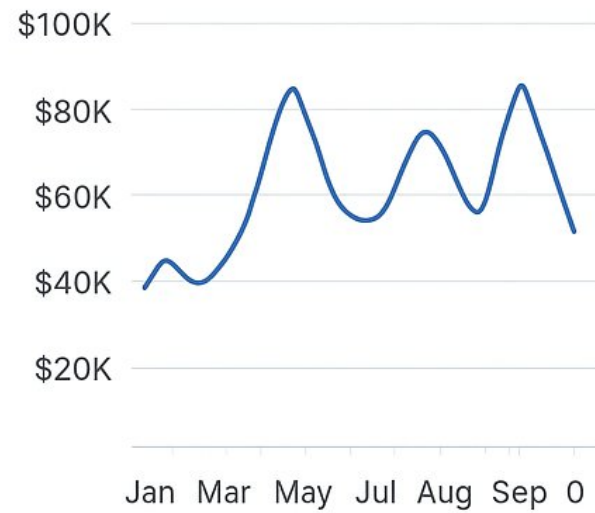
Sales by Region



Top Products



Sales Trends



Sales by Region

Category Time Location

Case Study 3: Smart Manufacturing Dashboard

Context:

A factory wants real-time monitoring of machine health and production efficiency.

Design Choices:

- **Machine status** → traffic-light color coding (Green = OK, Yellow = Warning, Red = Failure).
- **Production trend** → line charts for hourly output.
- **Downtime causes** → Pareto chart (sorted bar chart).
- **Alerts** → push notifications for threshold breaches.
- **Design Principles:**
  - Pre-attentive features (color, shape) for quick recognition.
  - Drill-down from factory-level → machine-level.

Impact:

Reduced downtime, predictive maintenance, higher production efficiency.

Smart Manufacturing Dashboard

ACTIVE MACHINES

28

OUTPUT RATE

120 units/hr

Downtime

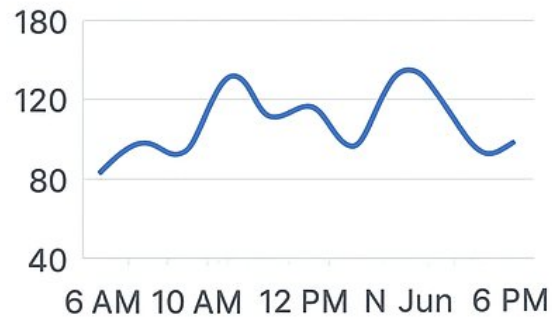
6 hr 20 min

6 hr 20 min

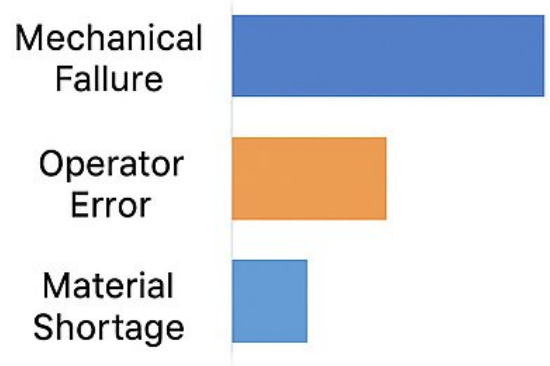
MACHINE STATUS

Machine	Status
M-01	OK
M-02	Warning
M-03	Failure
M-04	Failure

PRODUCTION TREND



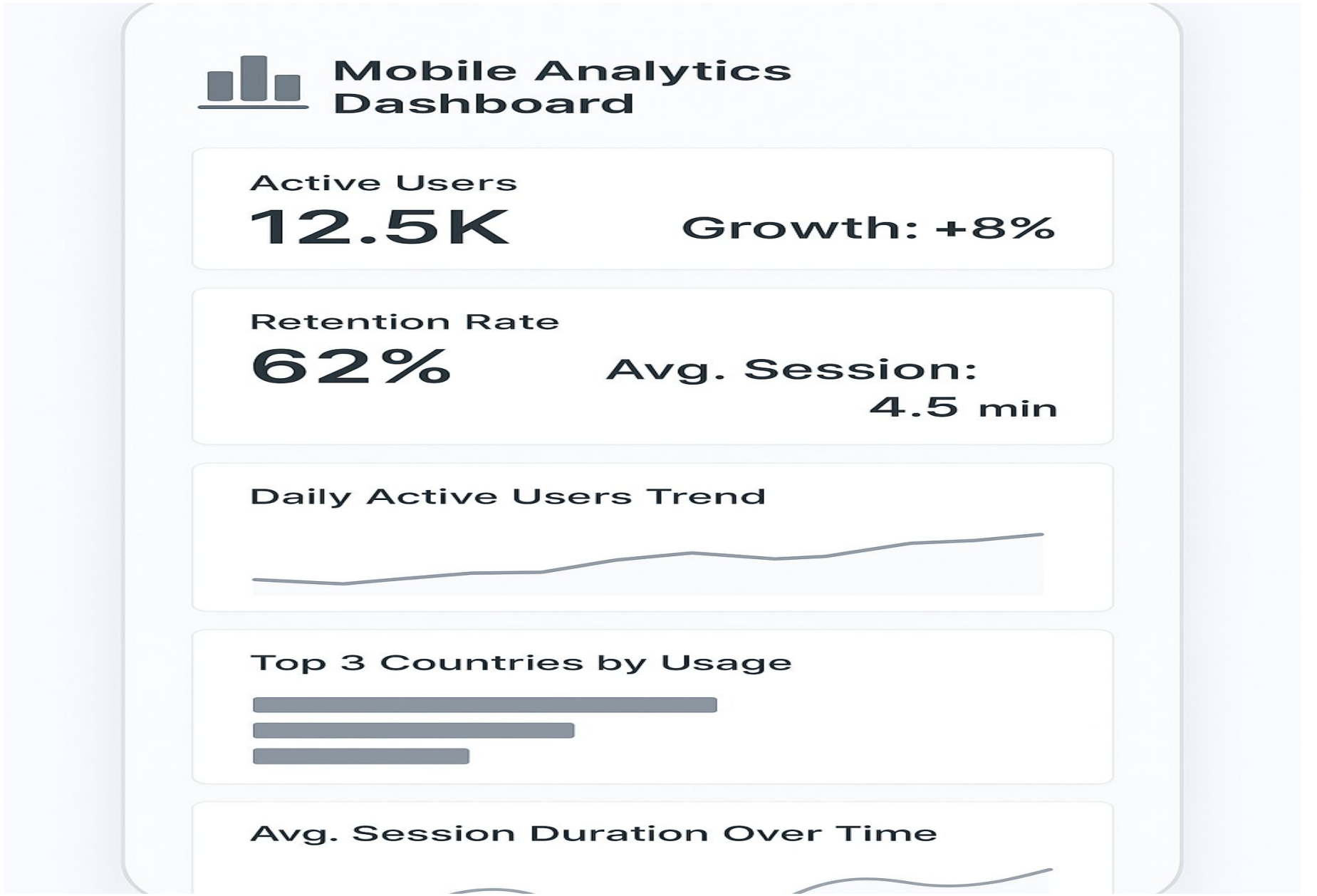
DOWNTIME CAUSES



ALERTS

Threshold breach: Line 2 output below 80 units/hr

. Apply visualization design to avoid clutter in mobile analytics dashboards.



Common Clutter Problems in Mobile Dashboards

- Too many KPIs crammed into one screen.
- Overuse of charts (bar, pie, line) without prioritization.
- Small text, poor spacing, inconsistent colors.
- No hierarchy → user doesn't know where to look first.

Visualization Design Strategies to Avoid Clutter

1. Prioritize & Reduce

- Show only **key KPIs** (e.g., DAU, MAU, Retention Rate) on the main screen.
- Move detailed metrics to **drill-down views**.

2. Apply Gestalt Principles

- **Proximity & Common Region:** Group related KPIs inside cards.
- **Similarity:** Use consistent colors/icons for related data (e.g., green for growth, red for drop).
- **Hierarchy:** Make the most important metric larger and central.

3. Use Progressive Disclosure

- Summary at first glance → tap to expand for details.
- Example: Show “Engagement Rate” at high level; tap card to view trend graph.

4. Minimalist Charts

- Prefer **sparklines or mini trend lines** over bulky charts.
- Avoid pie charts on mobile (hard to read); use bars/lines.

5. Responsive Layout

- One column scrolling layout (not grid overload).
- Each section should be easy to scan in **2–3 seconds**.

Apply the visualization pipeline to create a dashboard for student grades.

Visualization Pipeline Applied to Student Grades

Data Collection (Raw Data)

- Student information: Name, Roll No, Class, Subject.
- Grades: Marks in each subject, GPA, Pass/Fail status.
- Attendance % (optional for performance insight).

Raw data Table

Student	Math	Science	English	GPA	Attendance
Aditi	85	92	78	3.6	90%
Rohan	70	65	72	2.8	80%
Meera	95	88	91	3.9	96%

2.Data Transformation

- Calculate **Average per Subject**.
- Compute **Top 5 Students** by GPA.
- Categorize GPA ranges (High, Medium, Low).
- Attendance correlation with grades.

3. Visual Mapping

- Map data to appropriate chart types:
  - **Bar Chart** → Average score per subject.
  - **Pie Chart** → Grade distribution (A, B, C, D).
  - **Leaderboard (Table)** → Top 5 students with GPA.
  - **Line Chart** → Attendance vs GPA trend.

4. Rendering

- Choose colors: Green = High GPA, Yellow = Medium, Red = Low.
- Use clear labels, icons (🎓 for students, 📊 for performance).
- Place visuals in a grid layout for clarity.

5. **View (Final Dashboard for Users)**

