DA,
Spring, 2026

# Data Visualization

***Faculty of DS & AI***
***Spring semester, 2026***

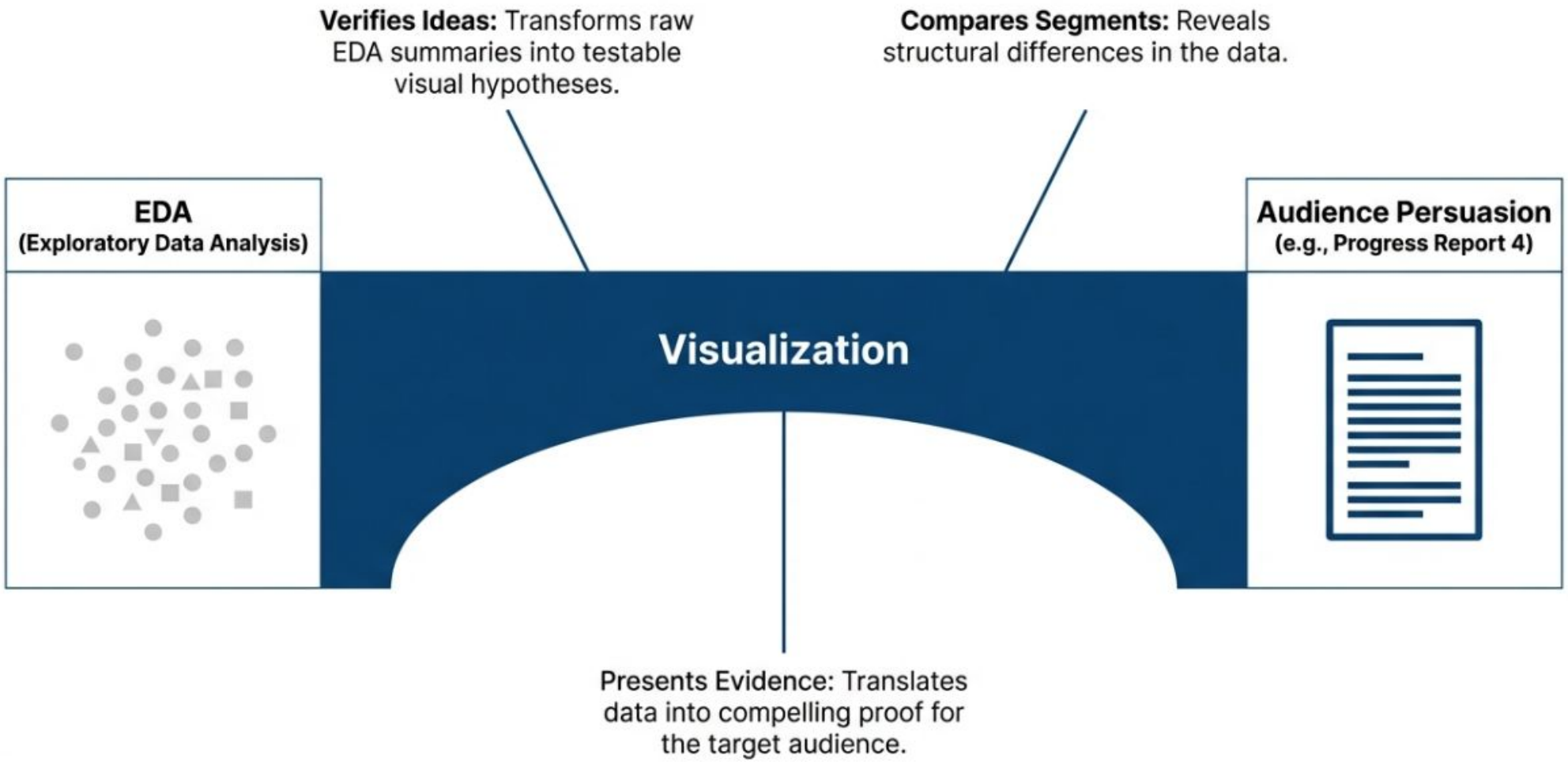
Trong-Nghia Nguyen

Business AI Lab

# Content

- **Visualization in the analytics workflow**
- Multivariate Visualization
- Design principles & common mistakes
- Dashboards in Excel
- Survival analysis

# Visualization in the analytics workflow



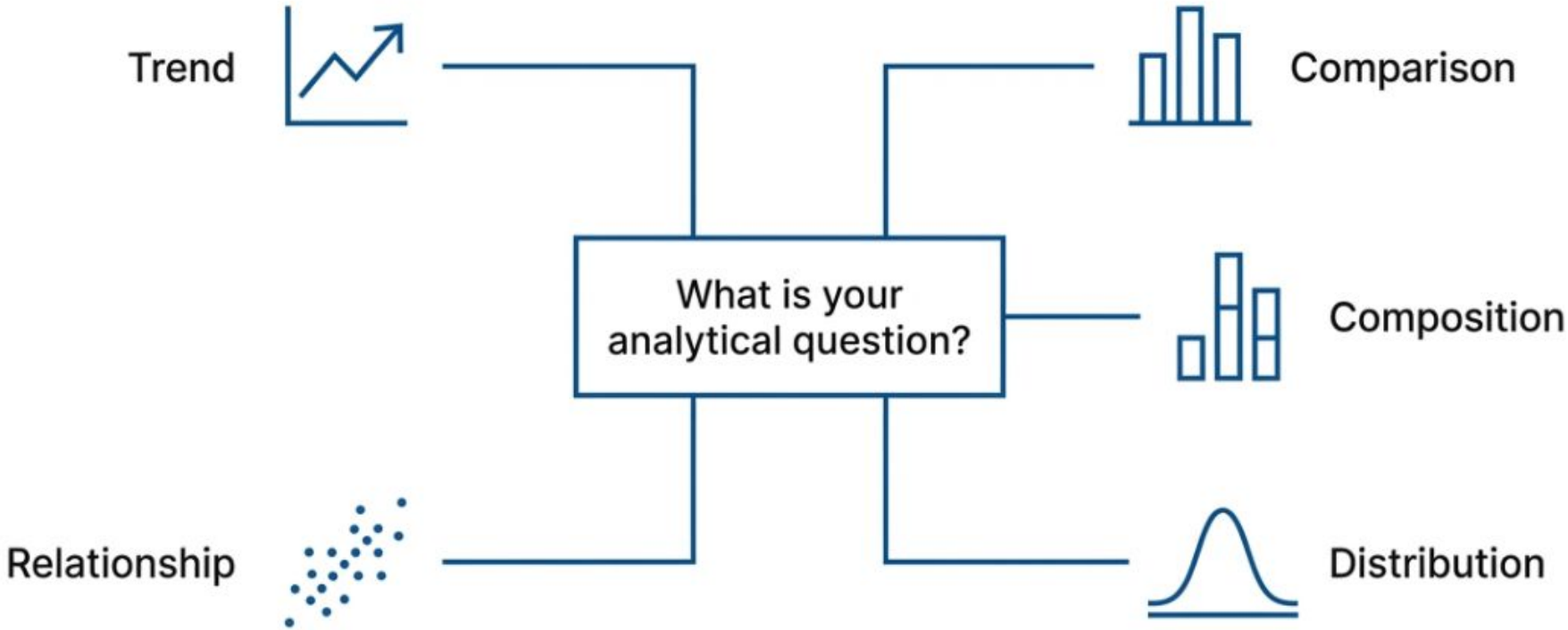
# Visualization in the analytics workflow

| | Exploratory | Explanatory |
|---|---|---|
| **Primary Goal** | Find hidden patterns, trends, anomalies. | Answer a specific question, deliver a core message. |
| **Target Audience** | The Analyst / Self. | External Stakeholders / Decision Makers. |
| **Phase in Process** | Early / EDA Phase. | Final / Communication Phase. |
| **Level of Detail** | High / Exhaustive / Granular. | Distilled / Highly Curated. |

# Visualization in the analytics workflow

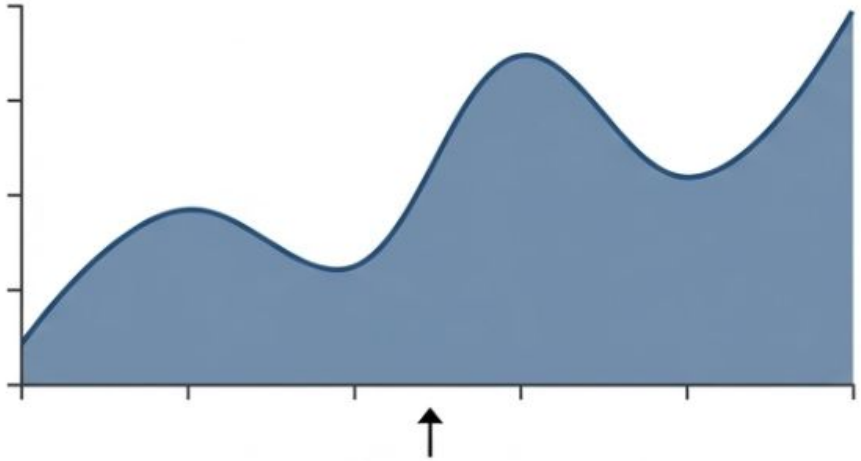
## The Cognitive Match

The core principle of data communication is matching your analytical question to the correct chart architecture. **Misalignment creates distortion.**



# Visualization in the analytics workflow

## Trend

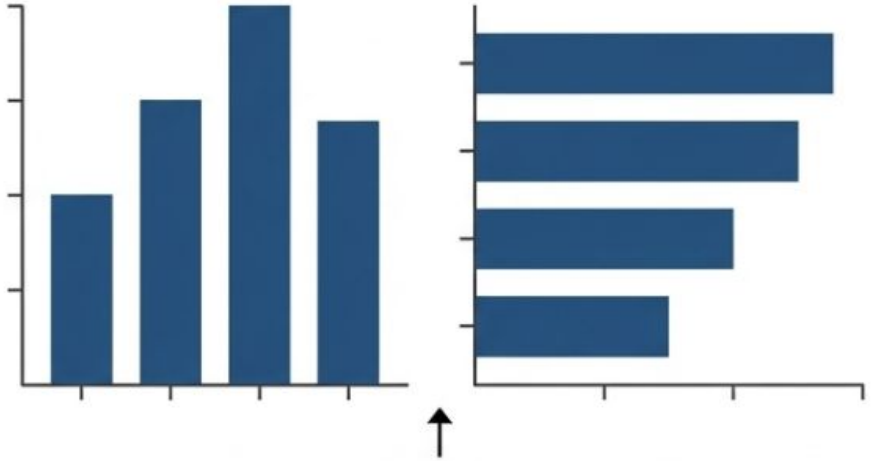


**Time (Continuous)**

Tracks variables over a continuous sequence.

**Example:** Tracking sales revenue month-over-month.

## Comparison



**Categories (Discrete)**

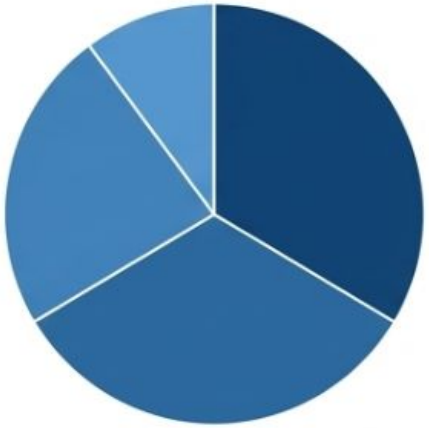
Evaluates magnitude across independent groups.

**Example:** Comparing total sales performance across geographical regions.

# Visualization in the analytics workflow

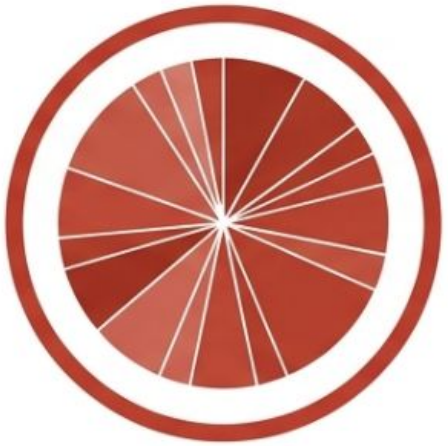
## Composition

### Part-to-Whole Constraint Graphic



**Clear Constraint (Scientifically Valid)**

Use only when categories are strictly limited and showing a part of the whole.

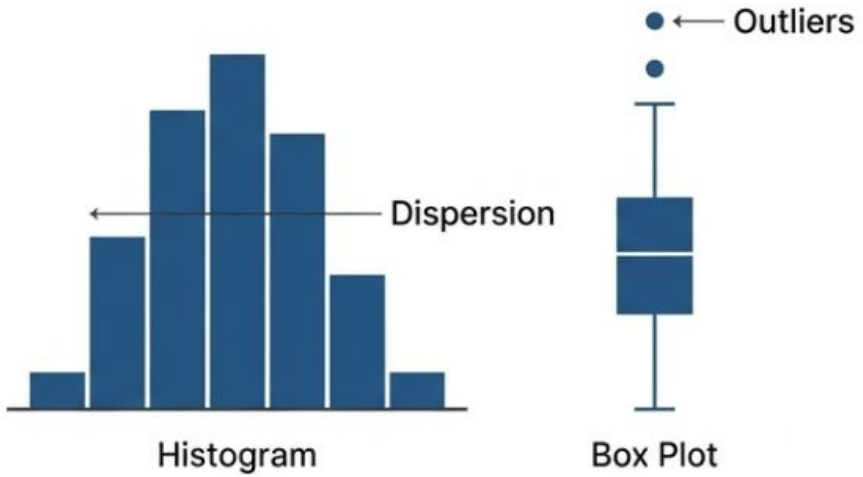


**Cognitive Clutter (Invalid)**

The human eye cannot accurately judge angles and areas of numerous small segments.

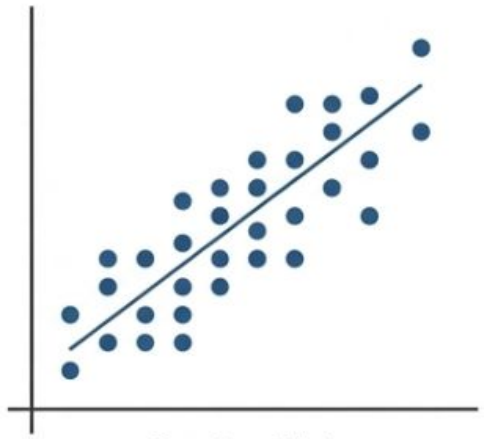
# Visualization in the analytics workflow

## Distribution




Reveals the spread, central tendency, and anomalies within a single variable.

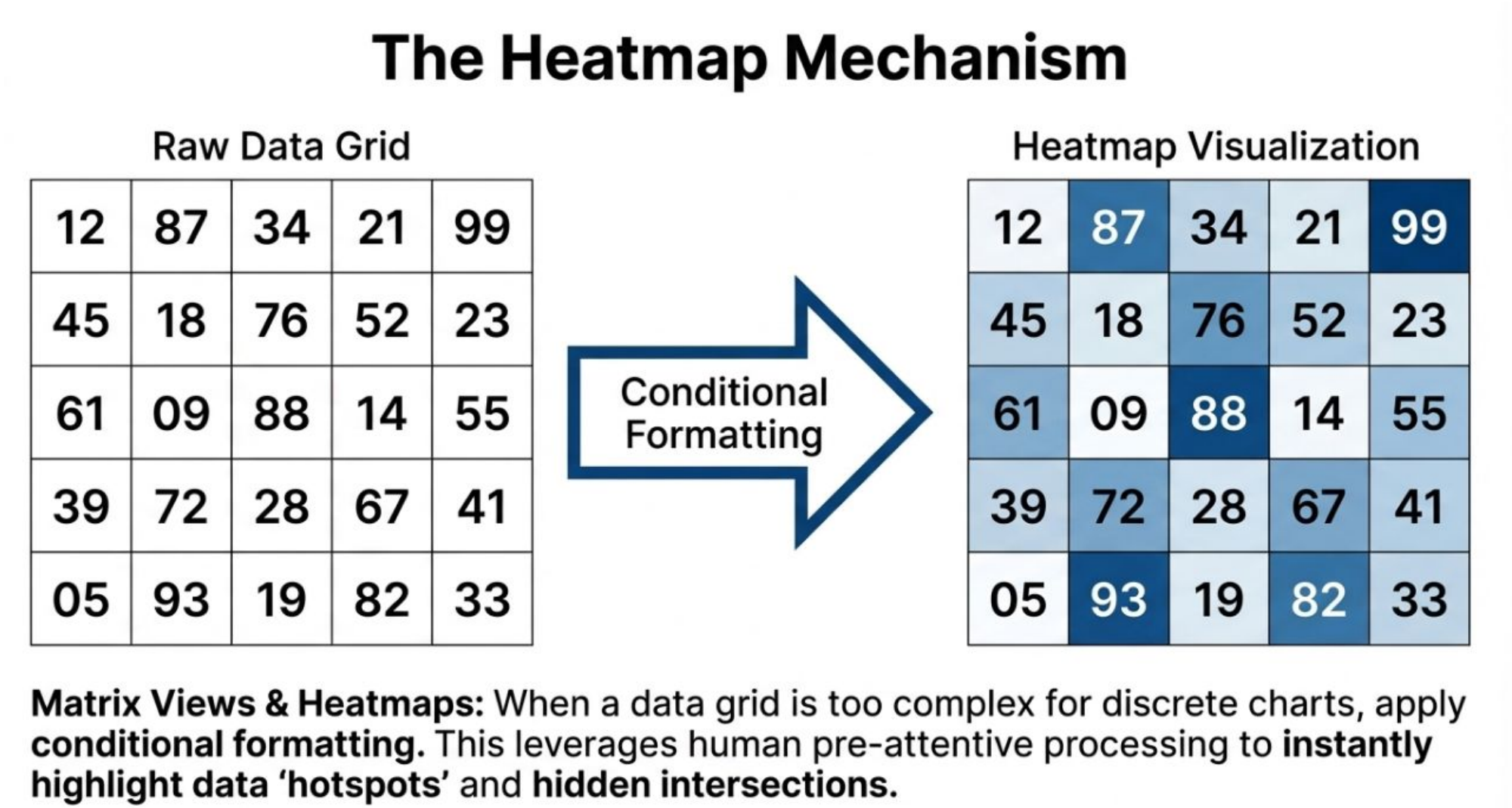
## Relationship




Examines the correlation between two continuous variables.

Example: Plotting the relationship between discount rates and total profit.

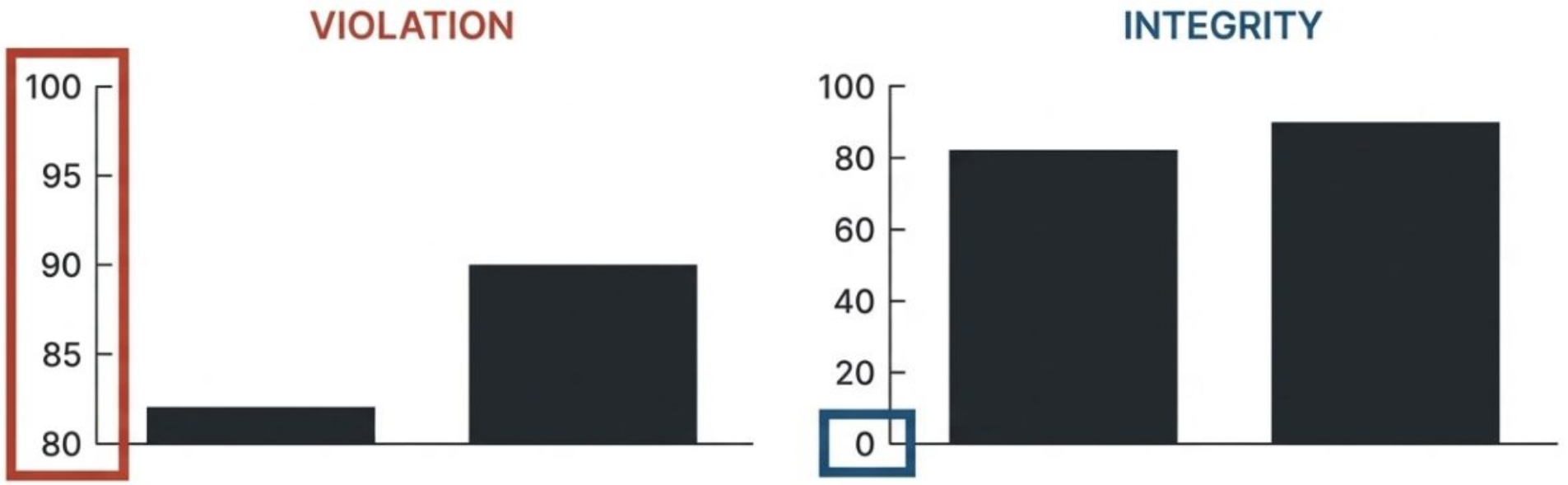
# Visualization in the analytics workflow




**Matrix Views & Heatmaps:** When a data grid is too complex for discrete charts, apply **conditional formatting.** This leverages human pre-attentive processing to **instantly highlight data 'hotspots'** and **hidden intersections.**

# Visualization in the analytics workflow

## Honest Encoding: The Baseline Mandate




**Distortion: Truncated Baseline**

Exaggerates minor variances.

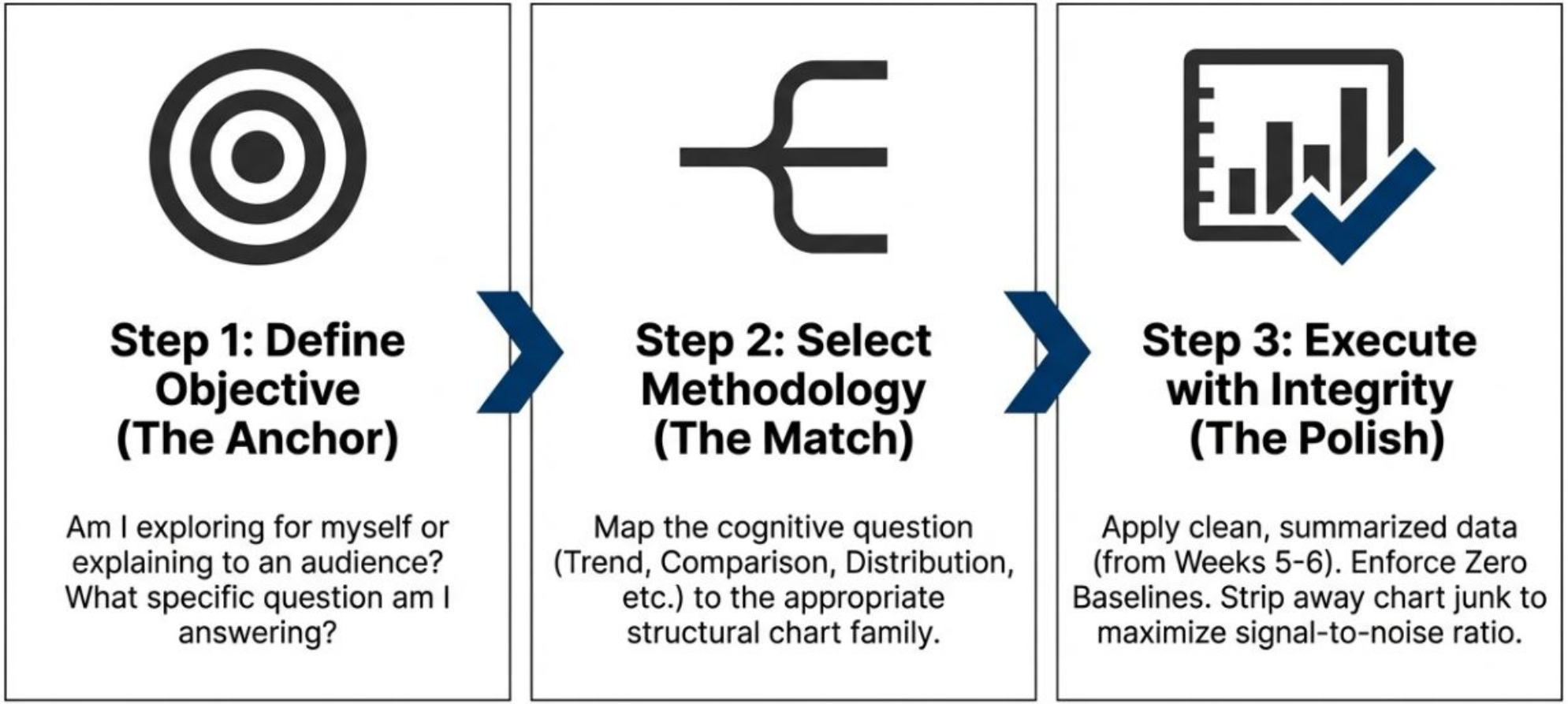
**Scientific Integrity: Zero Baseline**

Bar and Column charts must always begin at zero to ensure column length accurately reflects data proportionality.

# Visualization in the analytics workflow

The Student Practice Orientation: A 3-Step Protocol

## 3-Step Decision Flowchart

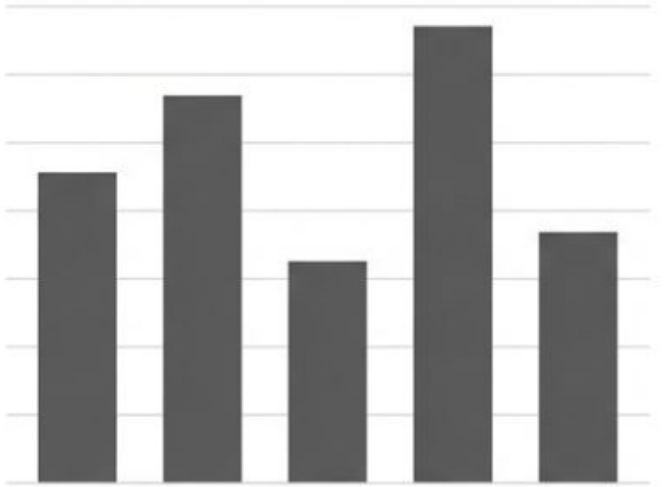


# Content

- Visualization in the analytics workflow
- **Multivariate Visualization**
- Design principles & common mistakes
- Dashboards in Excel
- Survival analysis

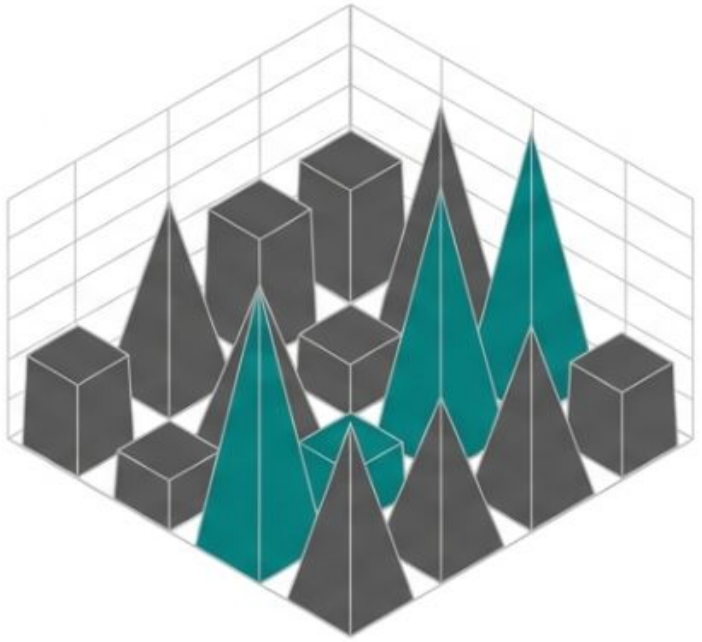
# Multivariate Visualization

Real-world questions demand multiple dimensions of data.



Single Variable Analysis

Basic metrics fail to capture intersecting realities.



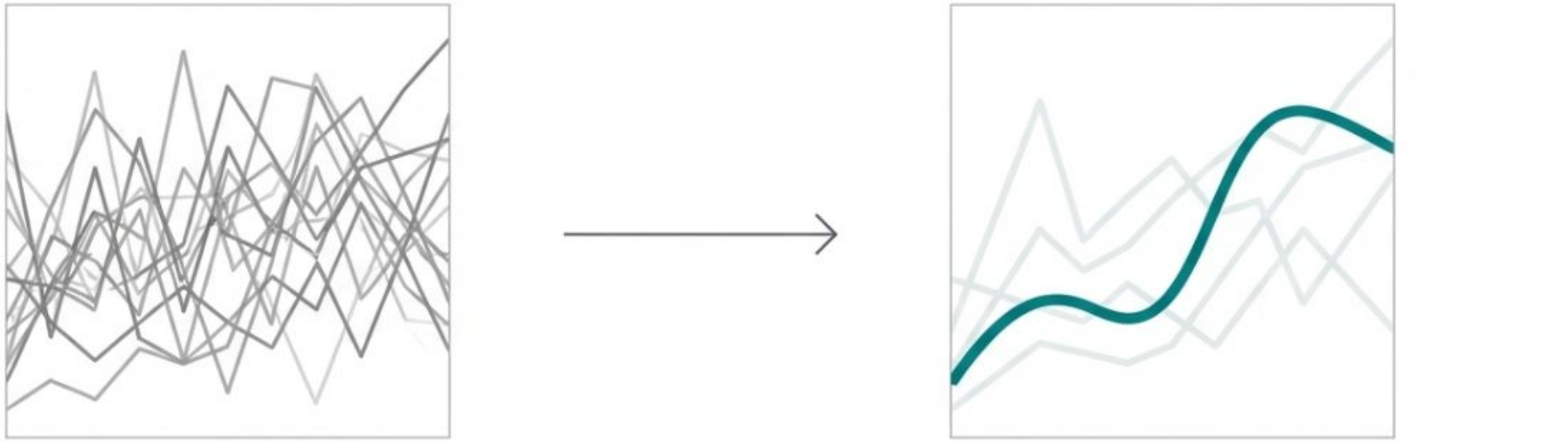
Multivariate Analysis

Allows simultaneous analysis of multiple dimensions to find complex relationships

How does Profit change by Region AND by Category?
Does the relationship between Discount and Profit differ by Customer Segment?

# Multivariate Visualization

The analytical challenge is balancing detail with clarity.



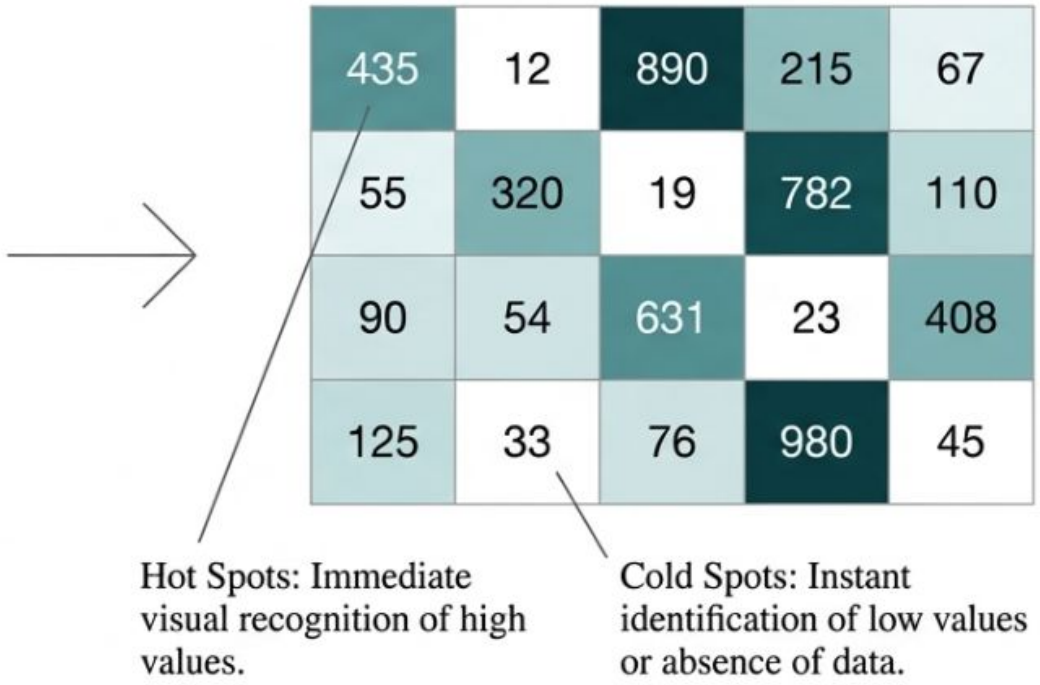

**The Objective:** Presenting more than one data dimension concurrently without sacrificing immediate interpretability.

**The Risk:** As dimensions increase, visual cognitive load grows exponentially.

# Multivariate Visualization

## Heatmaps reveal density and structure within complex grids.

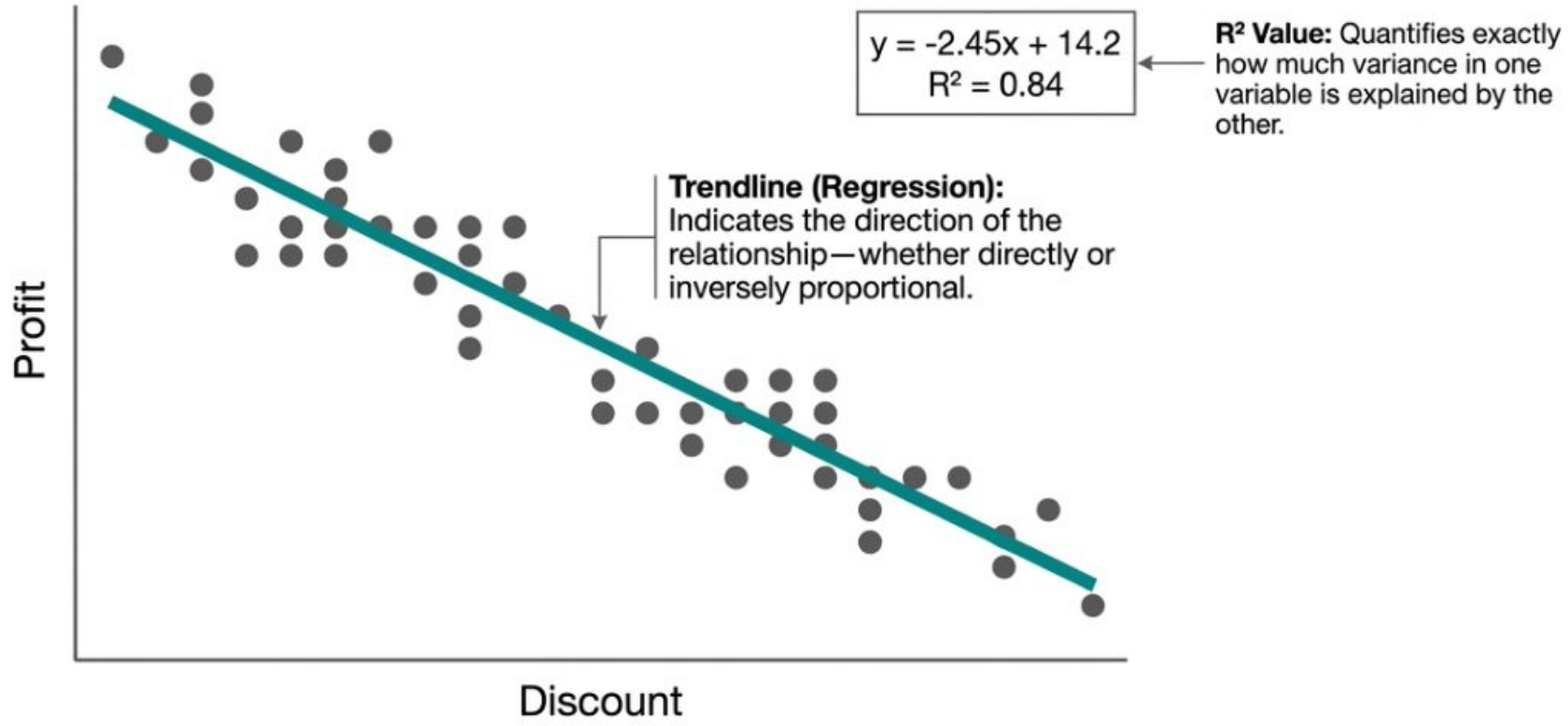
| | | | | |
|---|---|---|---|---|
| 435 | 12 | 890 | 215 | 67 |
| 55 | 320 | 19 | 782 | 110 |
| 90 | 54 | 631 | 23 | 408 |
| 125 | 33 | 76 | 980 | 45 |




Research Application: Mapping event density over time or visualizing correlation matrices between patient vital signs.

# Multivariate Visualization

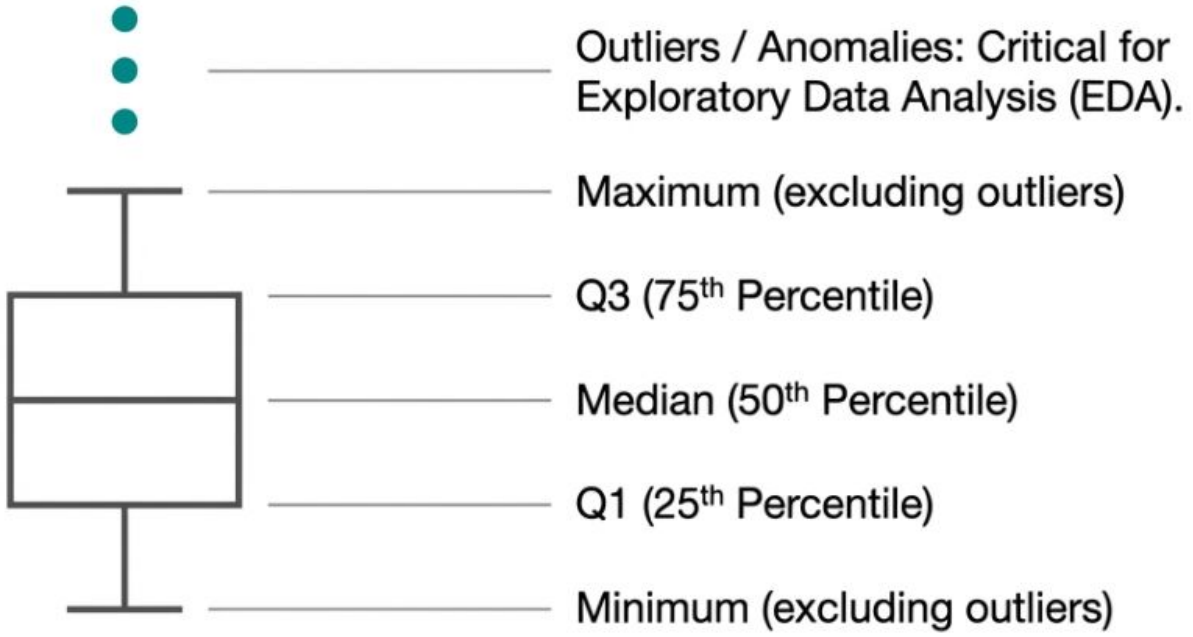
Scatter plots quantify relationships between continuous variables.




Medical Application: Tracing the exact correlation between administered medication dosage and patient reaction times.

# Multivariate Visualization

Box plots provide a scientific standard for comparing distributions.




Enables side-by-side comparative analysis of distinct segments (e.g., Heart rate variance in Group A vs. Group B) while instantly exposing anomalies that other charts obscure.

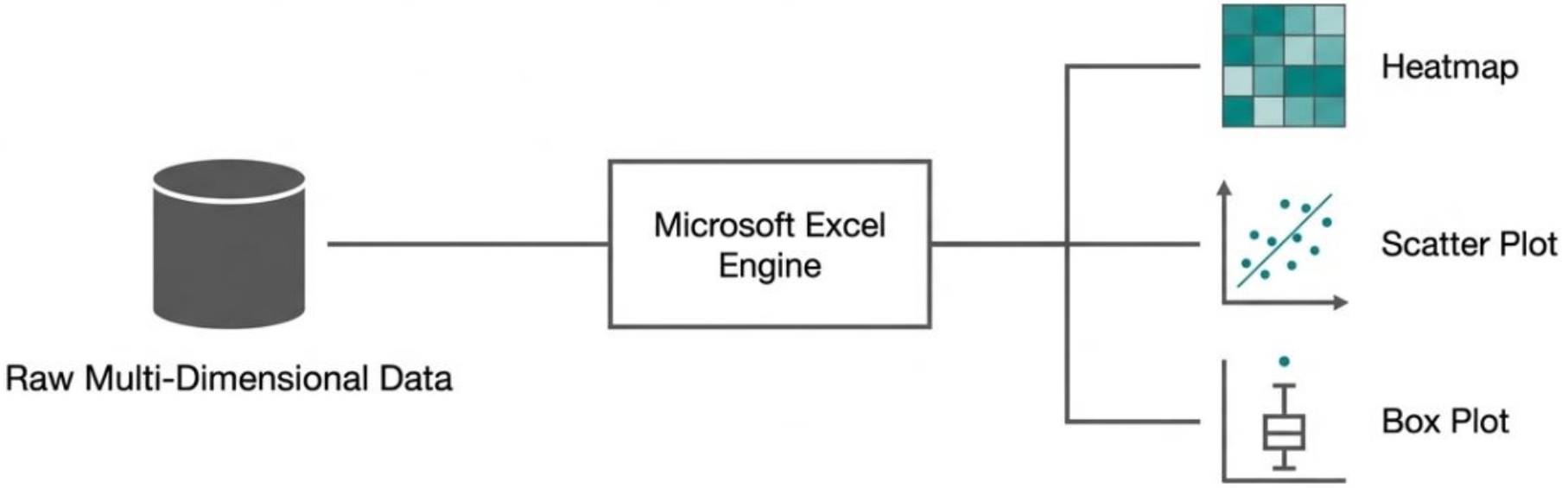
# Multivariate Visualization

Selecting the optimal visualization depends on data structure and analytical goals.

| Chart Type | Primary Goal | Ideal Data Structure | Key Strength |
|---|---|---|---|
| Heatmap | Identify Density & Patterns | Categorical × Categorical | Handles complex grid structures effortlessly. |
| Scatter Plot | Find Relationships & Correlation | Quantitative × Quantitative | The undisputed champion for continuous variables. |
| Box Plot | Compare Distributions & Outliers | Categorical × Quantitative | Unmatched for spotting anomalies in Exploratory Data Analysis. |

# Multivariate Visualization

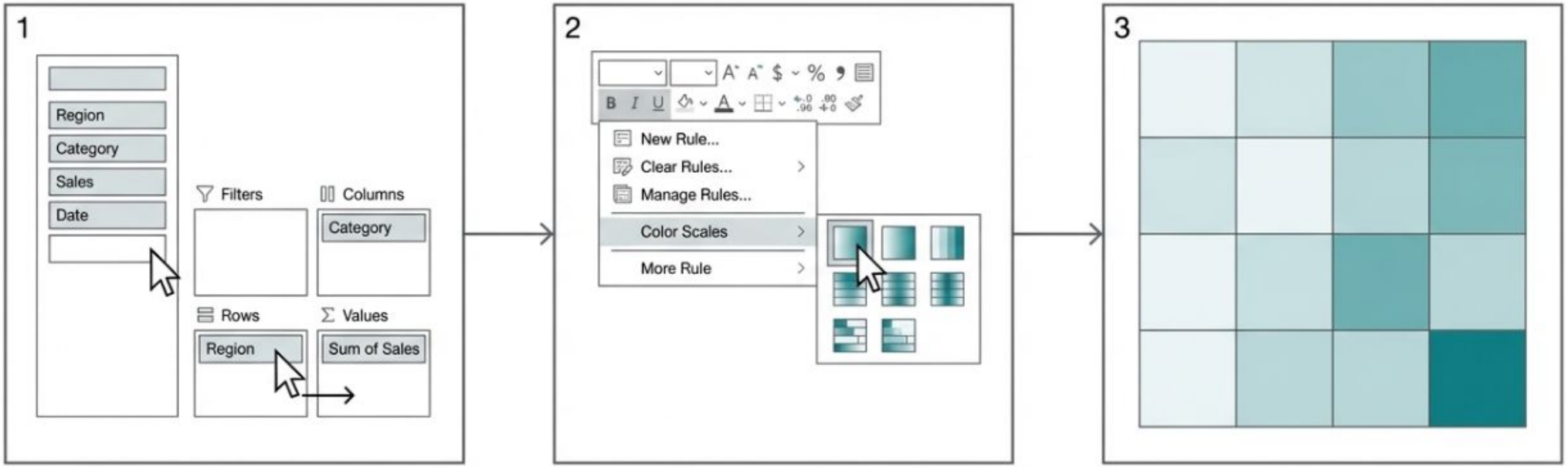
Executing multivariate visualizations requires mapping theory to software tools.




Transitioning from statistical theory to practical application. The following methodologies provide a standardized, step-by-step handbook for generating academic-grade visualizations natively within Excel.

# Multivariate Visualization

## Generating Heatmaps from dense data grids grids in Microsoft Excel.



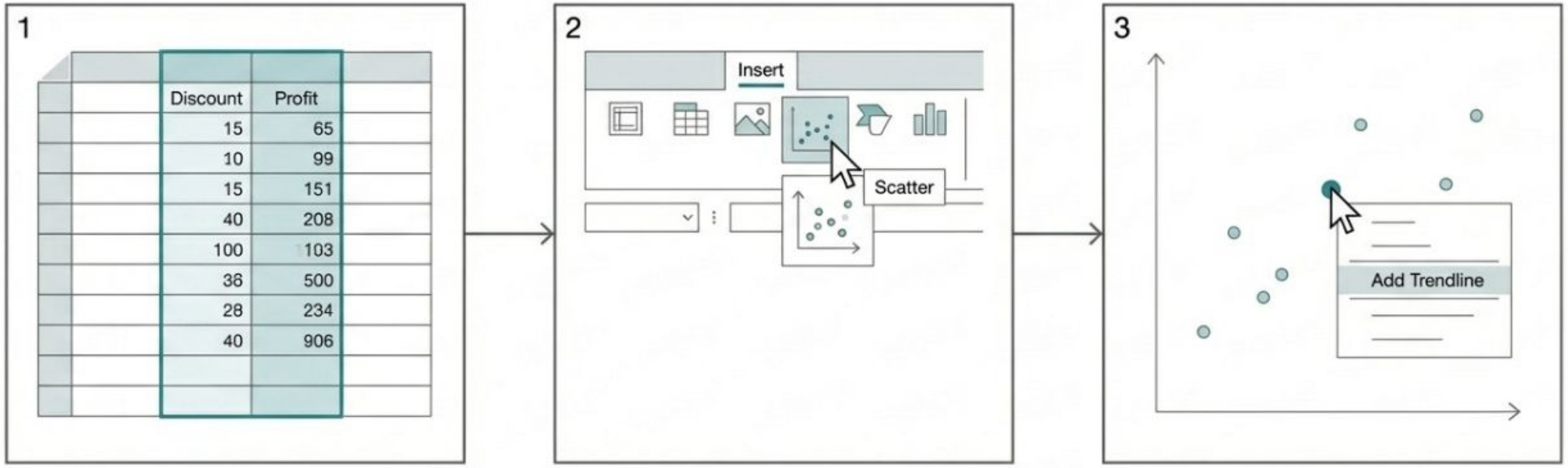

Build PivotTable: Assign Region to Rows, Category to Columns, and Sum of Sales to Values.

Highlight numeric range -> Home tab -> Conditional Formatting -> Color Scales.

Instant visual density grid.

# Multivariate Visualization

## Generating Scatter Plots with statistical trendlines in Excel.



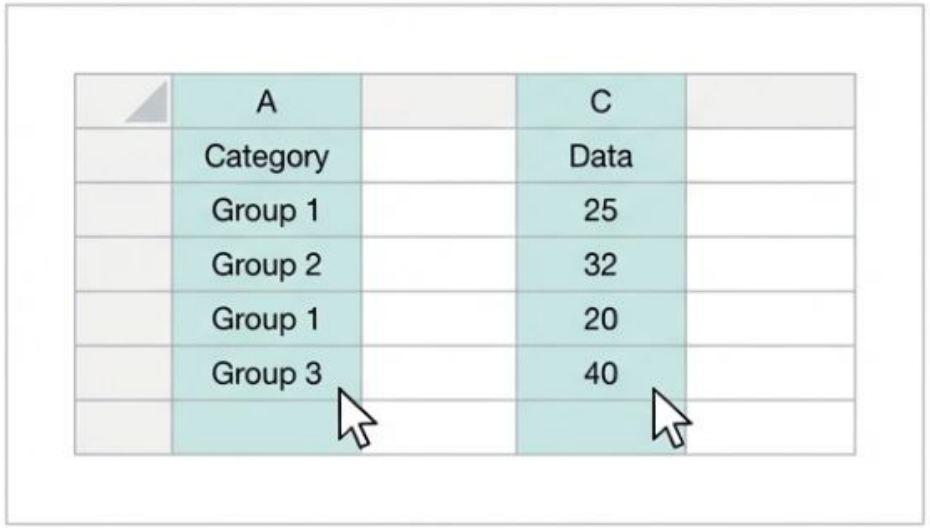

Select two corresponding quantitative columns (e.g., Discount and Profit).

Navigate to Insert tab -> Select Scatter icon.

Right-click data point -> Add Trendline -> Check "Display Equation on chart".

# Multivariate Visualization

## Generating Box Plots for categorical distribution distribution analysis in Excel.



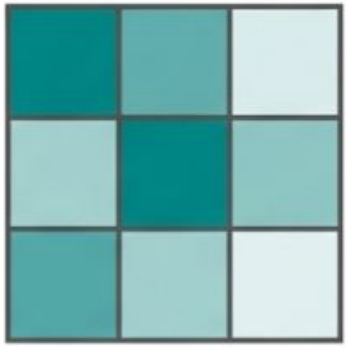

Select one categorical column and its corresponding numeric column simultaneously.


Navigate to Insert tab -> Statistical Chart dropdown -> Select Box and Whisker.

Pro-tip: Excel will automatically calculate Q1, Median, Q3, and plot outliers based on statistical standard deviations.

# Multivariate Visualization

Three distinct tools for uncovering multidimensional patterns.



Translates complex categorical grids into instant visual density maps using conditional color scaling.

Defines and quantifies the exact correlation and directional trend between two continuous variables.

Provides the scientific standard for comparative distribution analysis and immediate anomaly detection.

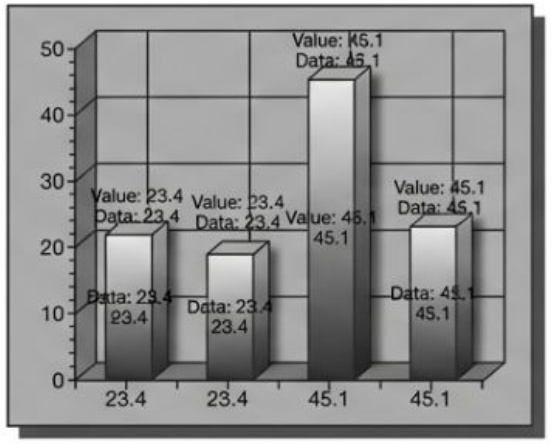
# Content

- Visualization in the analytics workflow
- Multivariate Visualization
- **Design principles & common mistakes**
- Dashboards in Excel
- Survival analysis

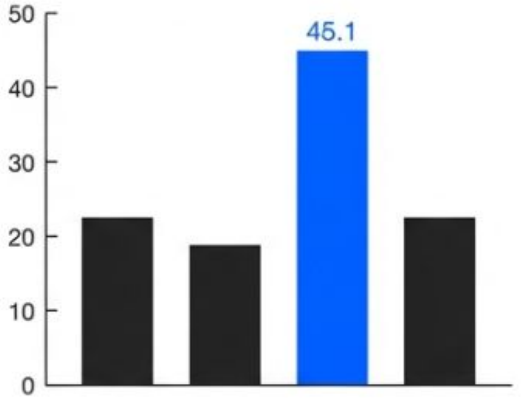
# Design principles & common mistakes

## Maximizing Signal by Minimizing Cognitive Load

[Raw Data] — [Chart Junk] = [Pure Signal]

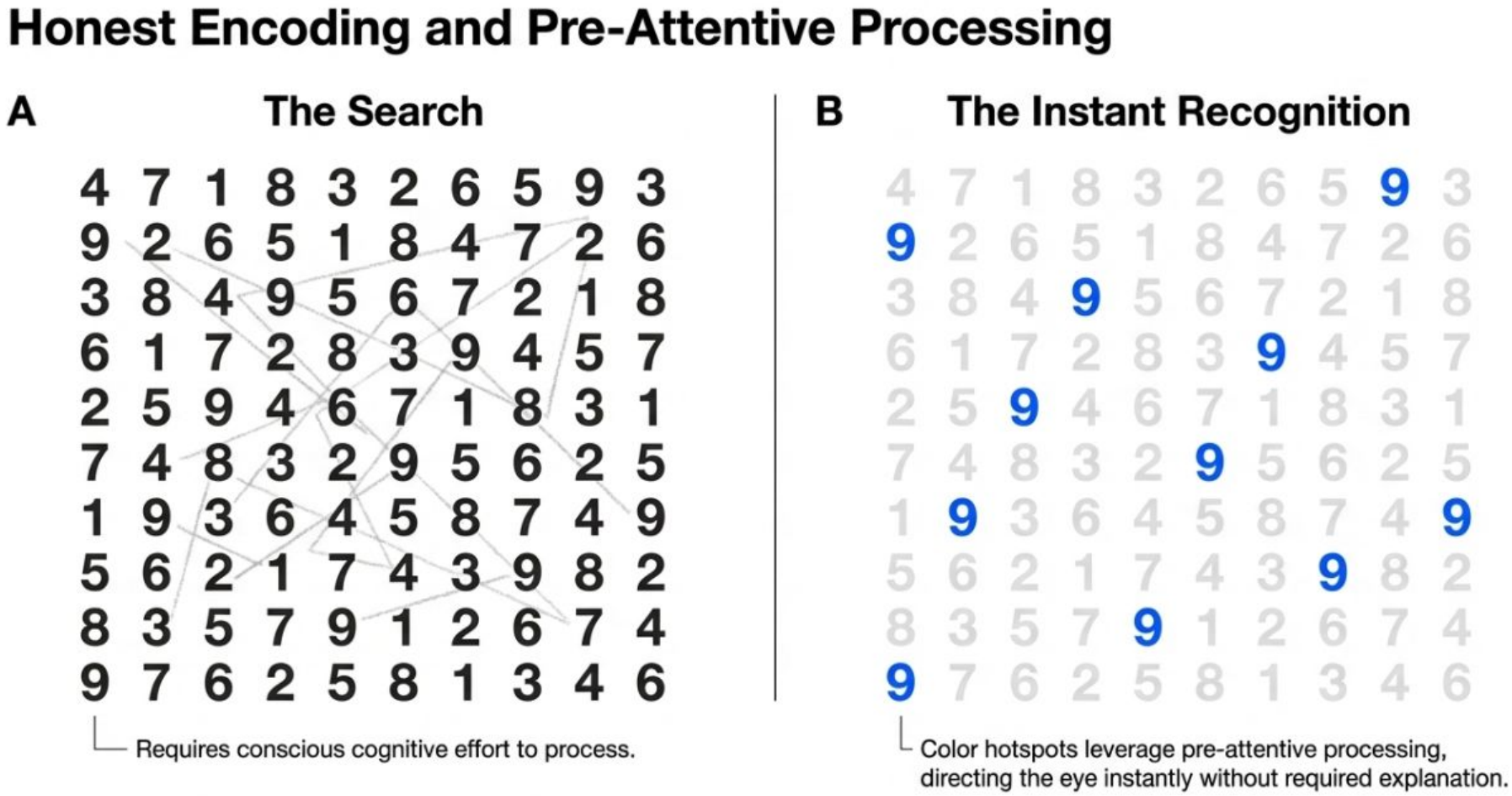



Heavy gridlines, 3D effects, redundant labels.




Cognitive load increases exponentially with each added data dimension. The ultimate goal of visualization is to balance necessary detail with absolute clarity.

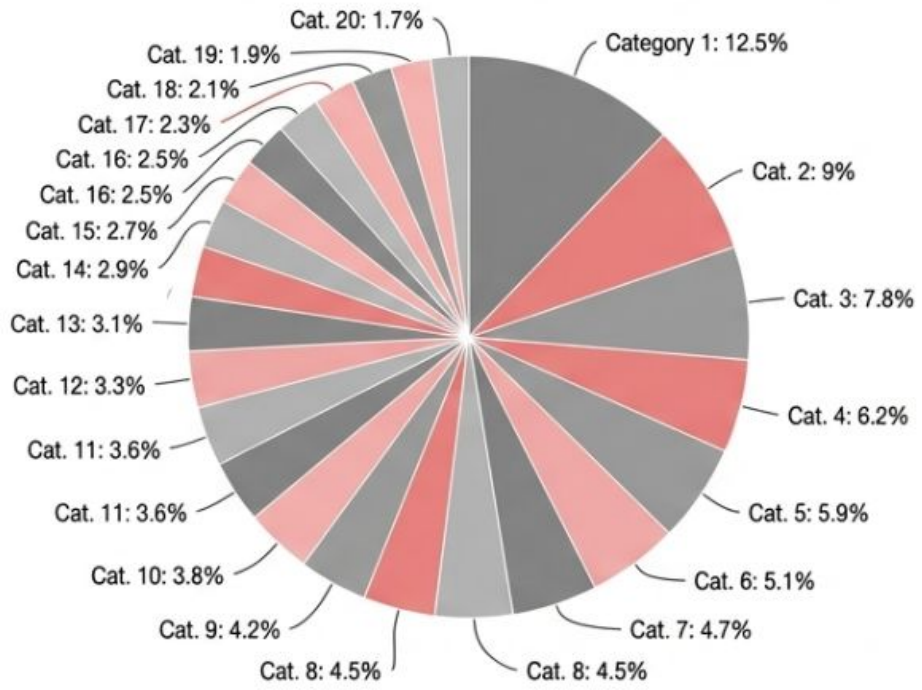
# Design principles & common mistakes



# Design principles & common mistakes

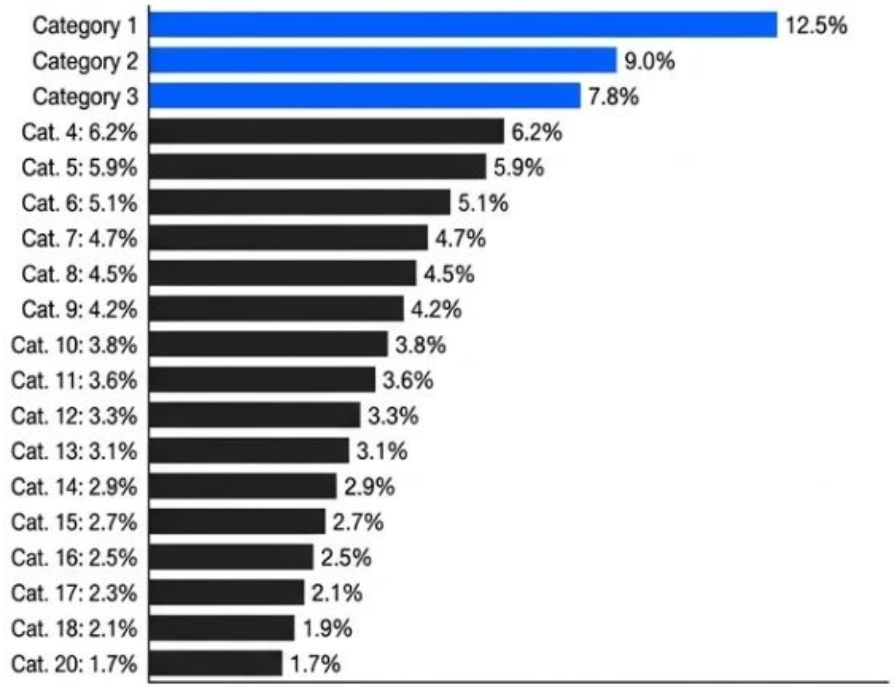
## Diagnostic Teardown: The Pie Chart Trap

A




**Clinical Callout**
**Cognitive Clutter:** The human eye cannot accurately judge angles and areas across multiple micro-segments.

B



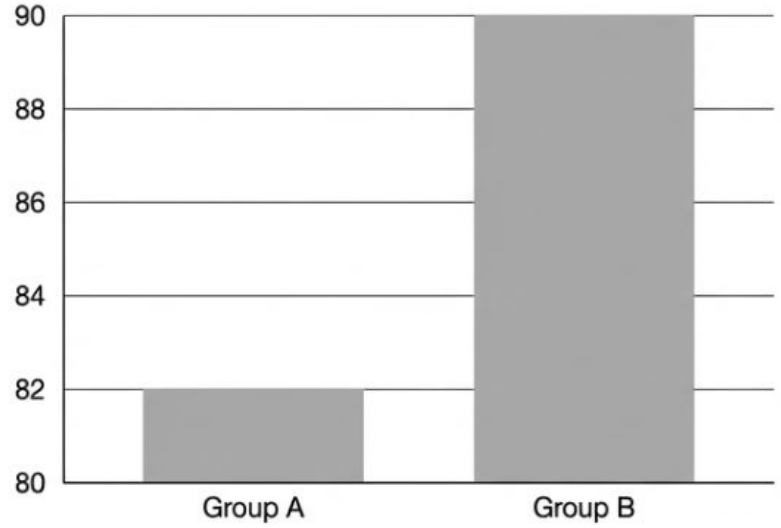

**Clinical Callout**
**Clear Constraint:** Limit pie charts exclusively to simple 'part-to-whole' relationships with minimal categories. Use sorted bar charts for multi-category ranking.

# Design principles & common mistakes

## Diagnostic Teardown: The Medical Distortion

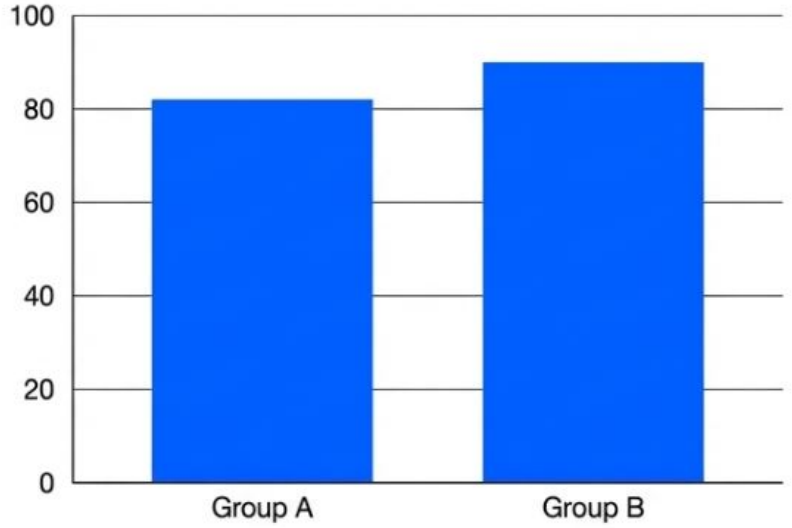
**Scenario:** Comparing heart rates of Group A (82 bpm) vs. Group B (90 bpm).

**A**




**Clinical Callout**
Truncated Baseline: Exaggerates minor variances. Visually implies a 500% difference, violating scientific integrity.

**B**

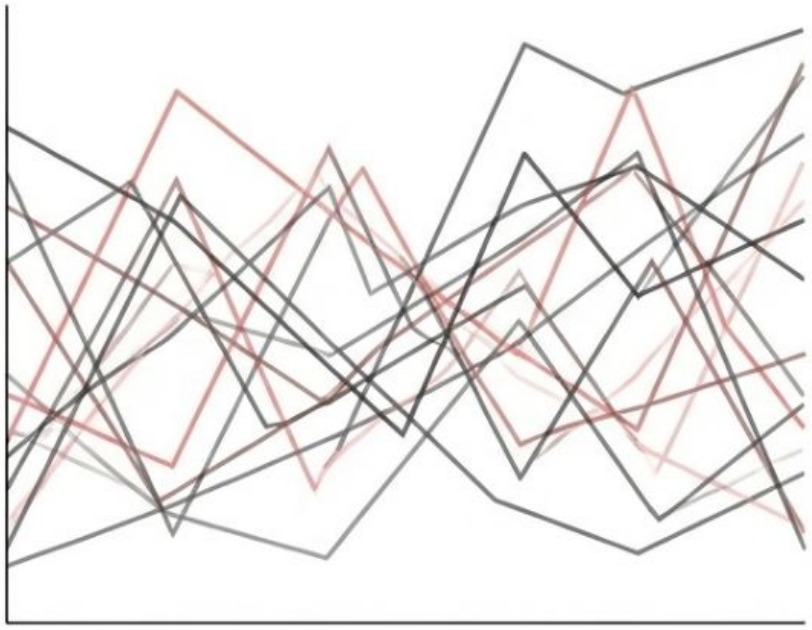



**Clinical Callout**
Zero Baseline Mandate: Bar and column charts must always anchor at zero to ensure column length accurately reflects true data proportions.

# Design principles & common mistakes

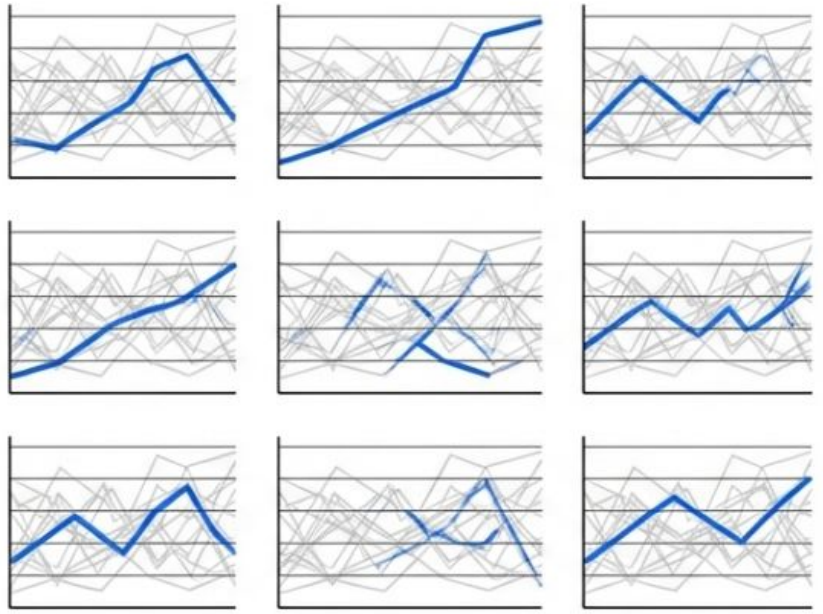
## Diagnostic Teardown: Information Overload

A



**Clinical Callout**
Information Overload: Multiple overlapping trends destroy individual pattern recognition.

B



**Clinical Callout**
Isolate the Signal: Use 'Small Multiples' grids or spatial Heatmaps to separate dimensions and restore clarity.

# Content

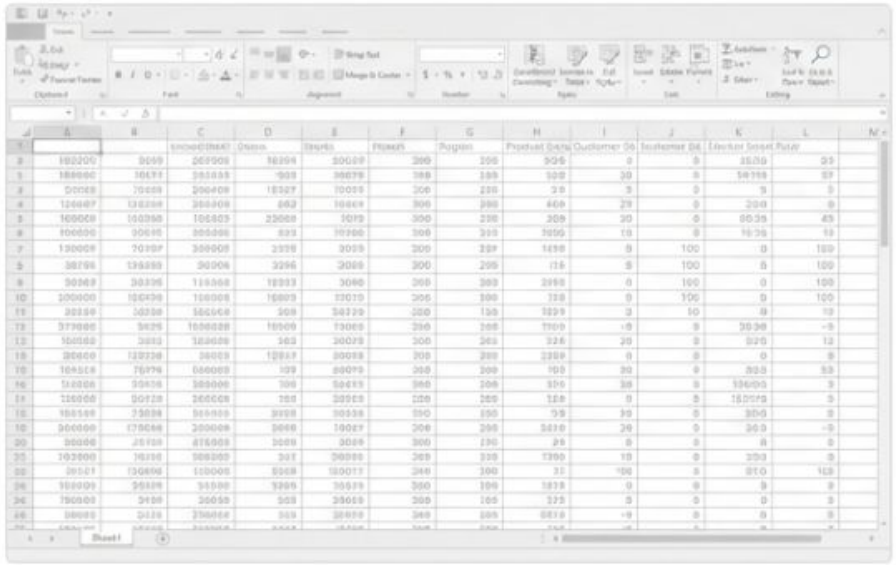
- Visualization in the analytics workflow
- Multivariate Visualization
- Design principles & common mistakes
- **Dashboards in Excel**
- Survival analysis (intro + Excel-friendly life-table workflow)

# Dashboards in Excel

## Dashboards are real-time, explanatory decision-making tools.

Dashboards provide a comprehensive view of system health to support rapid decision-making for external stakeholders or management. They allow users to filter and drill down into data without altering the underlying spreadsheet structure.

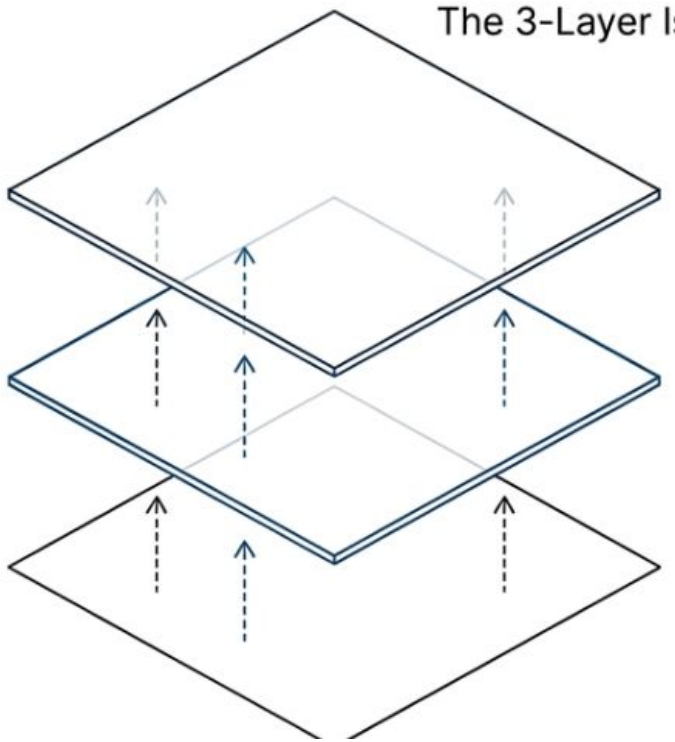
### The Old Paradigm



### The New Paradigm

# Dashboards in Excel

**Sustainable dashboards strictly separate data, logic, and display.**



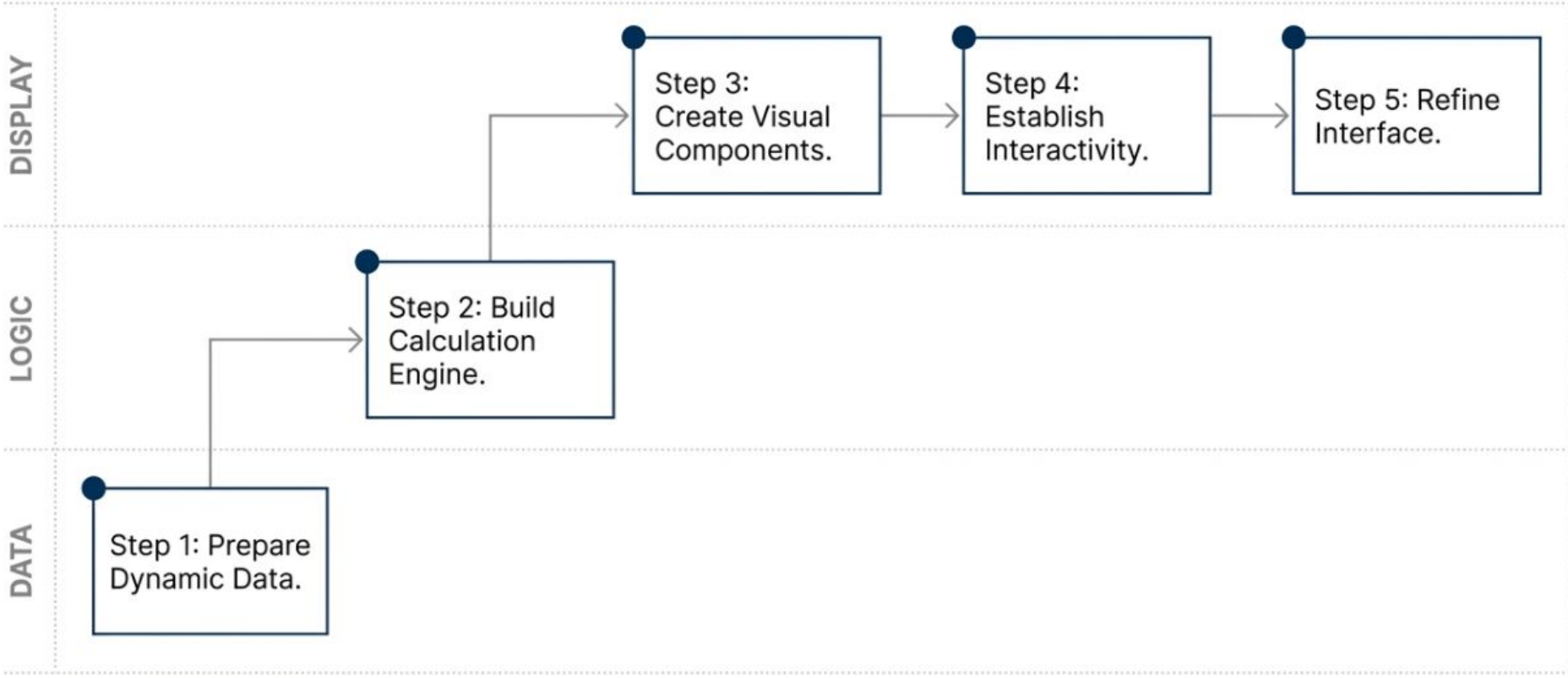

L3. Interface: Pivot Charts and Slicers communicating with the user.

L2. Analysis: Pivot Tables acting as calculation engines for multivariate summaries.

L1. Raw Data: Cleaned, flat tables.

# Dashboards in Excel

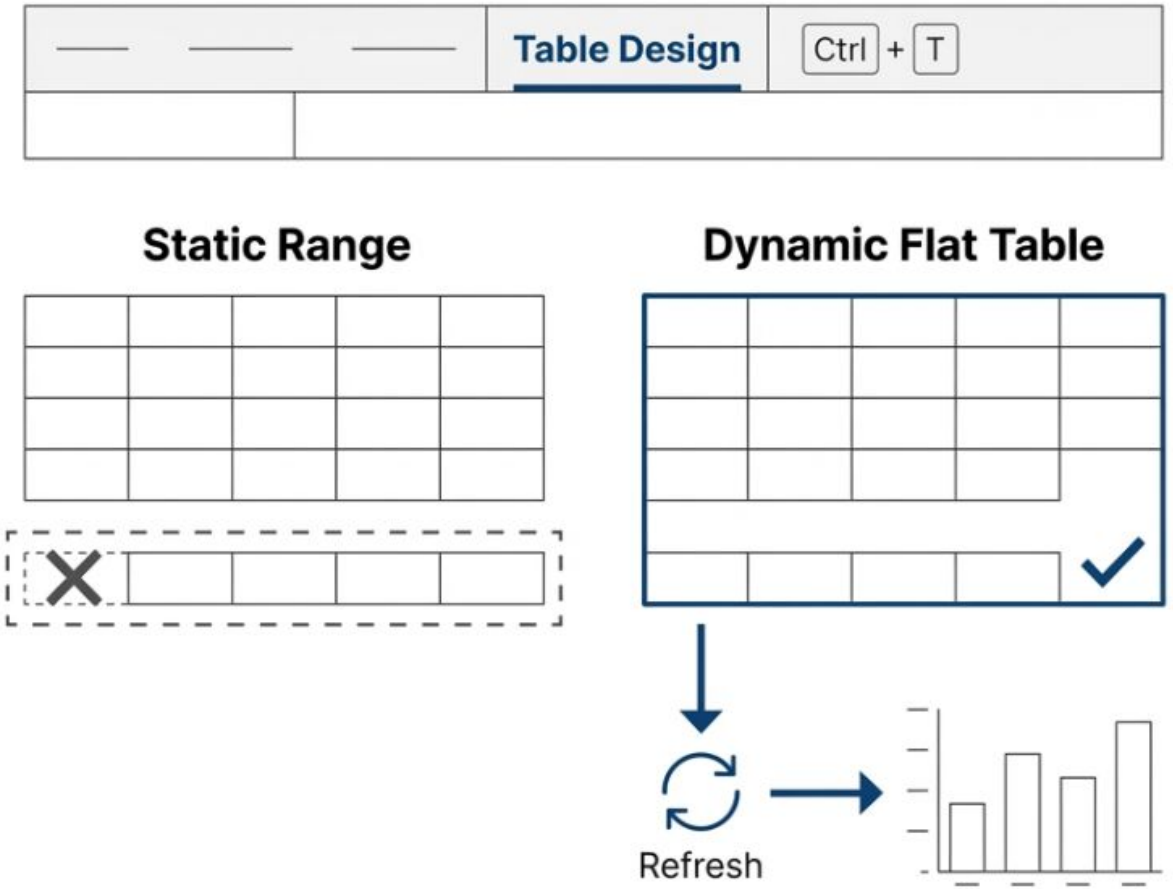
**The 5-Step Process maps directly to the underlying engine.**



# Dashboards in Excel

## Step 1: Convert raw data into dynamic flat tables

Using Table Format ensures that when new data is added, all connected dashboard charts update automatically without requiring the user to reselect data ranges.



# Dashboards in Excel

## Step 2: Isolate calculations by assigning one Pivot Table per analytical question.

**Action:** Insert -> PivotTable.
**Rule:** Keep calculation engines on a separate, dedicated sheet. This ensures the main dashboard interface remains uncluttered and functionally distinct.

# Dashboards in Excel

## Step 3: Match the visual output directly to the analytical intent.

**The Chart Selection Matrix**

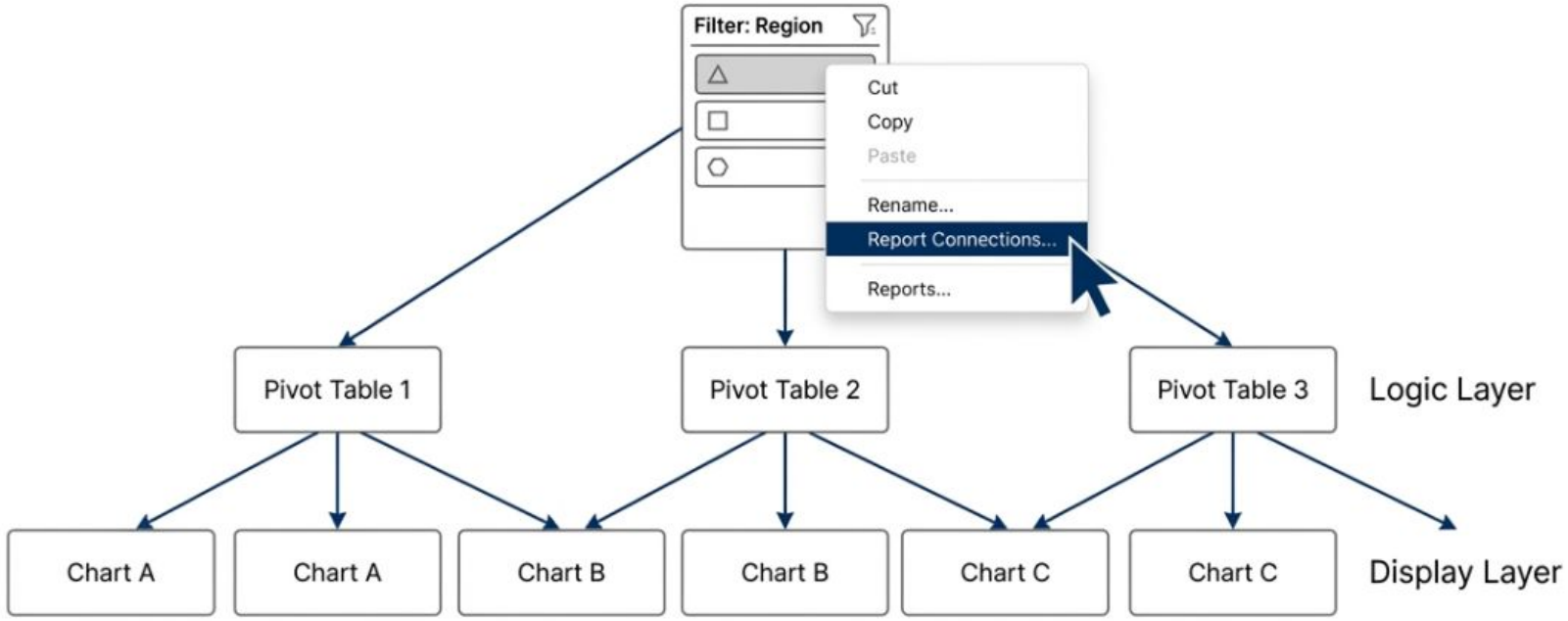
| Analytical Goal | Recommended Chart | Visual Structure | Excel Action |
|---|---|---|---|
| 1. Display temporal trends. | Line Chart. |  | PivotTable Analyze -> Pivot Chart |
| 2. Compare categorical sizes. | Bar / Column Chart. |  | PivotTable Analyze -> Pivot Chart |
| 3. Uncover quantitative relationships. | Scatter Plot. |  | PivotTable Analyze -> Pivot Chart |

# Dashboards in Excel

## Step 4: Centralize interactivity by synchronizing Slicer connections.

Clicking one Slicer button **must** update the entire dashboard simultaneously. This requires manually checking the linked Pivot Tables in the Report Connections menu.

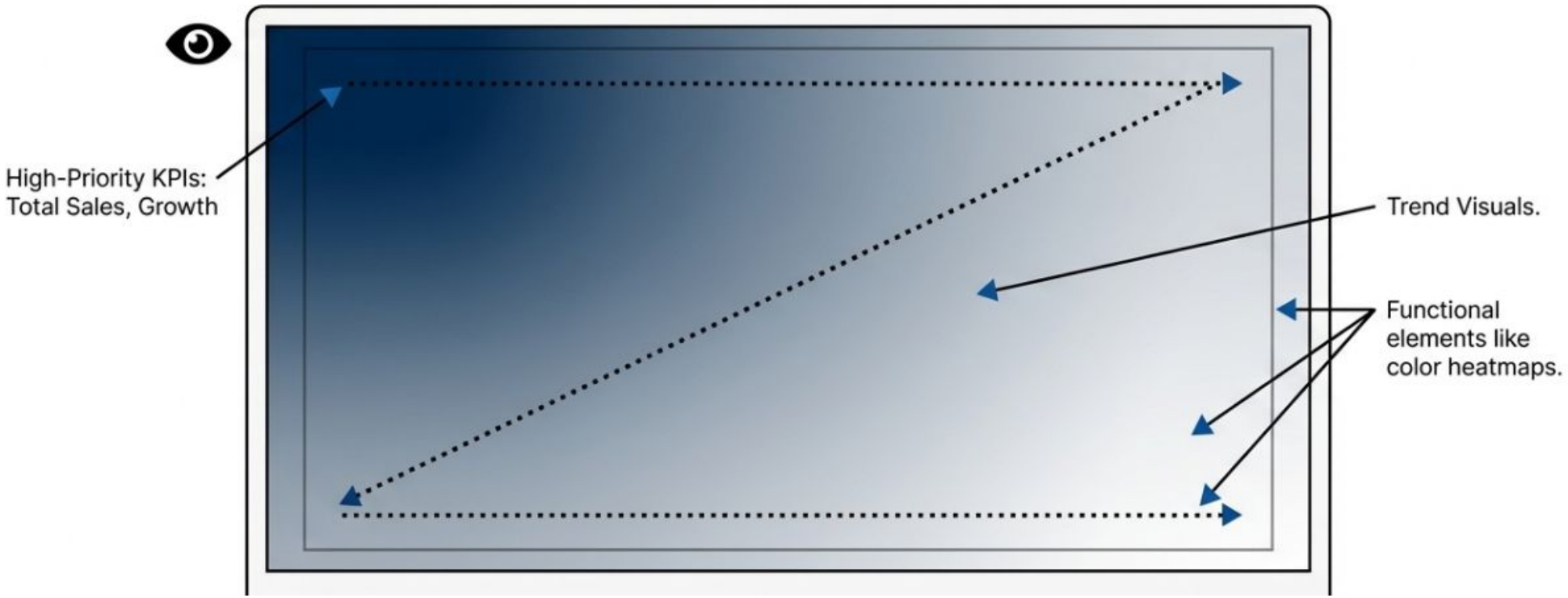
**The Synchronized Slicer Model**



# Dashboards in Excel

## Step 5: Prioritize visual hierarchy to satisfy the 5-Second Rule

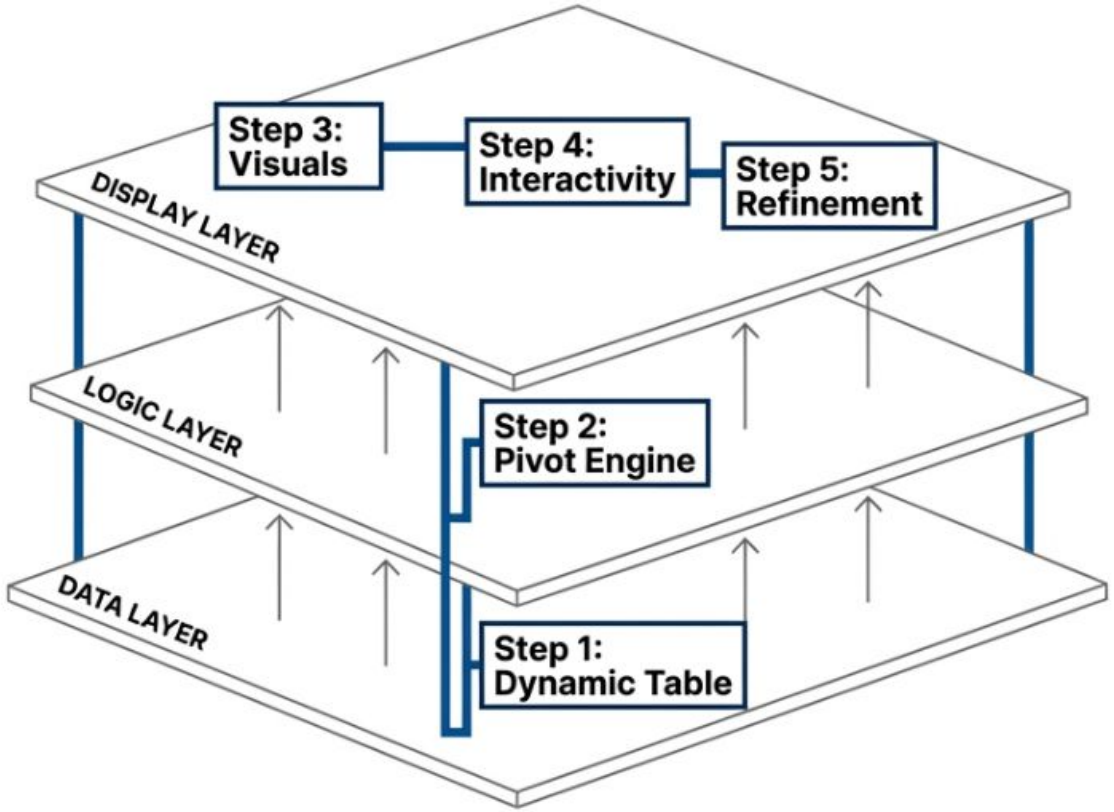
The viewer must understand the primary data trend within 5 seconds. Place the most critical metrics top-left. Use color functionally rather than decoratively.



# Dashboards in Excel

## Elegant visual design is the byproduct of rigorrous data architecture.

A dashboard only fails when the layers are compromised. A broken chart is usually a symptom of fragile data or unlinked logic.



# Content

- Visualization in the analytics workflow
- Multivariate Visualization
- Design principles & common mistakes
- Dashboards in Excel
- **Survival analysis**

# Survival analysis

## The Paradigm Shift: Why Standard Math Fails

### Standard Analysis

- Focuses only on IF an event happens (binary).
- Fails when subjects have different follow-up lengths or incomplete observations.
- Calculating a simple arithmetic mean causes severe bias.

### Survival Analysis


- Focuses on WHEN an event happens (**Time-to-Event**).
- Perfectly handles variable observation windows and dropouts.
- Preserves **partial data** without skewing baseline metrics.

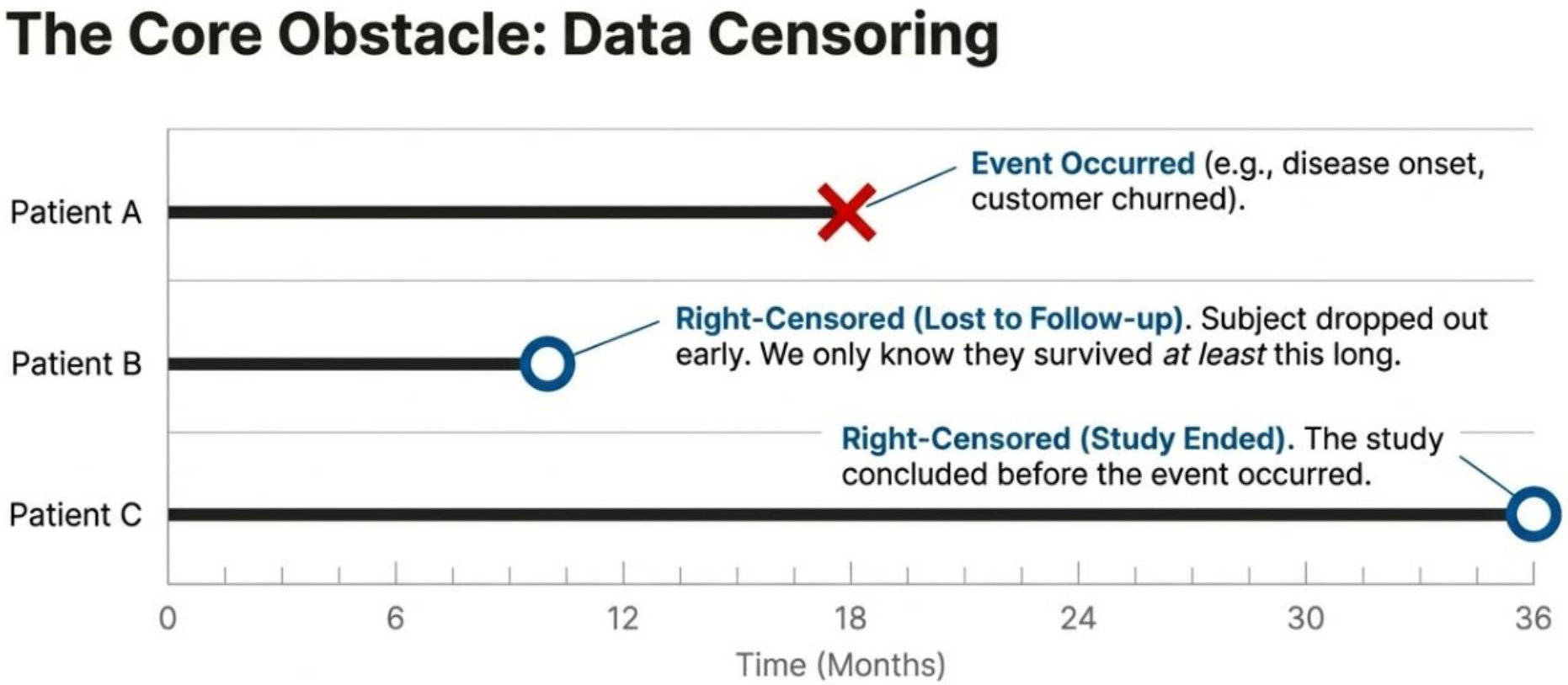
Real-World Applications:

Alzheimer's Onset

Machinery Failure

Customer Churn

# Survival analysis



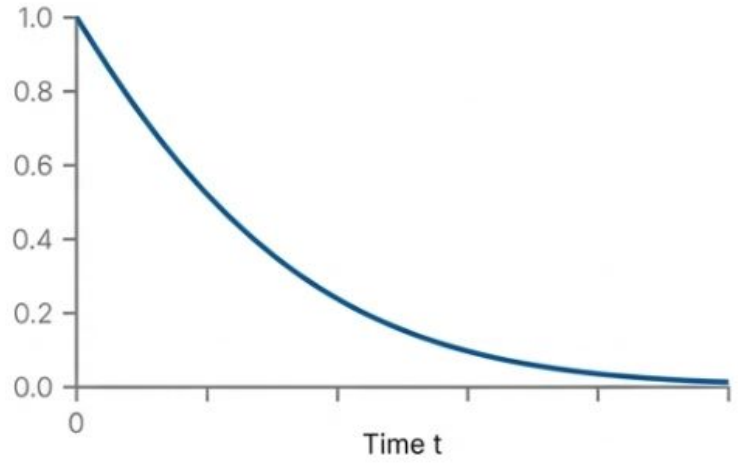

**Warning:** Ignoring censored data discards highly valuable partial information and **heavily skews baseline averages. Survival Analysis** mathematics uniquely solve this.

# Survival analysis

## Two Pillars of Survival Math

### The Survival Function S(t)

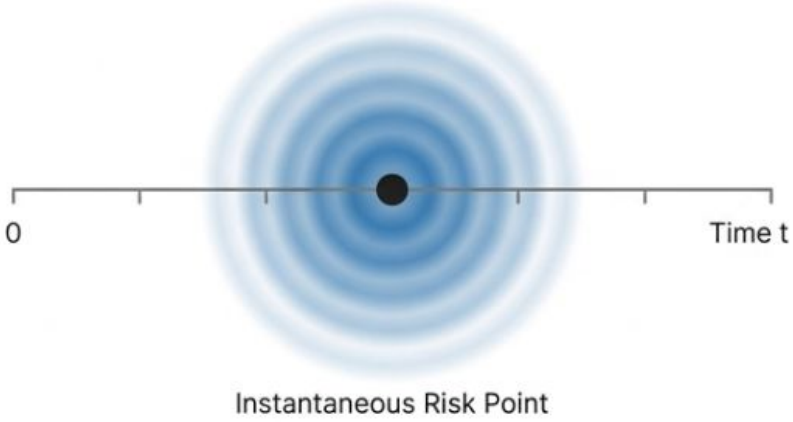
$$S(t) = P(T > t)$$




**Definition:** The cumulative probability that a subject will NOT experience the event beyond time t.

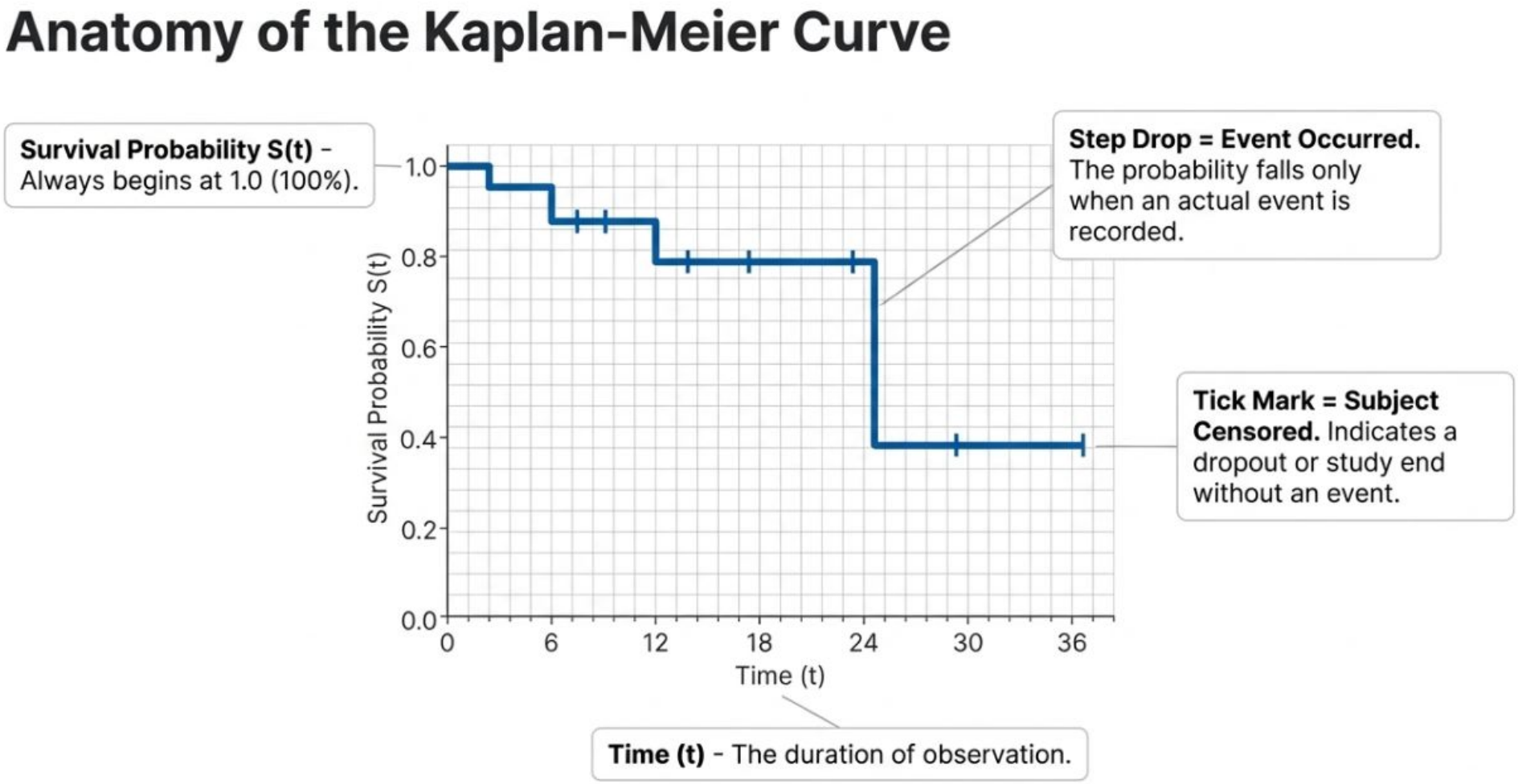
### The Hazard Function h(t)

$$h(t)$$




**Definition:** The instantaneous risk. The probability of the event happening *right now*, given the subject has survived up to this exact moment.

# Survival analysis



# Survival analysis

## Pre-Analysis: Understanding the Cohort

### PivotTable: Cohort Stratification

| Treatment Arm | Gender | Count | Mean Severity by Arm |
|---|---|---|---|
| Placebo | Male | 142 | 7.4 |
| Placebo | Female | 138 | 7.6 |
| Active Drug | Male | 150 | 4.2 |
| Active Drug | Female | 145 | 4.1 |

### Pre-Analysis Protocol

**Step 1:** Describe cohort distribution using categorical variables (Covariates).

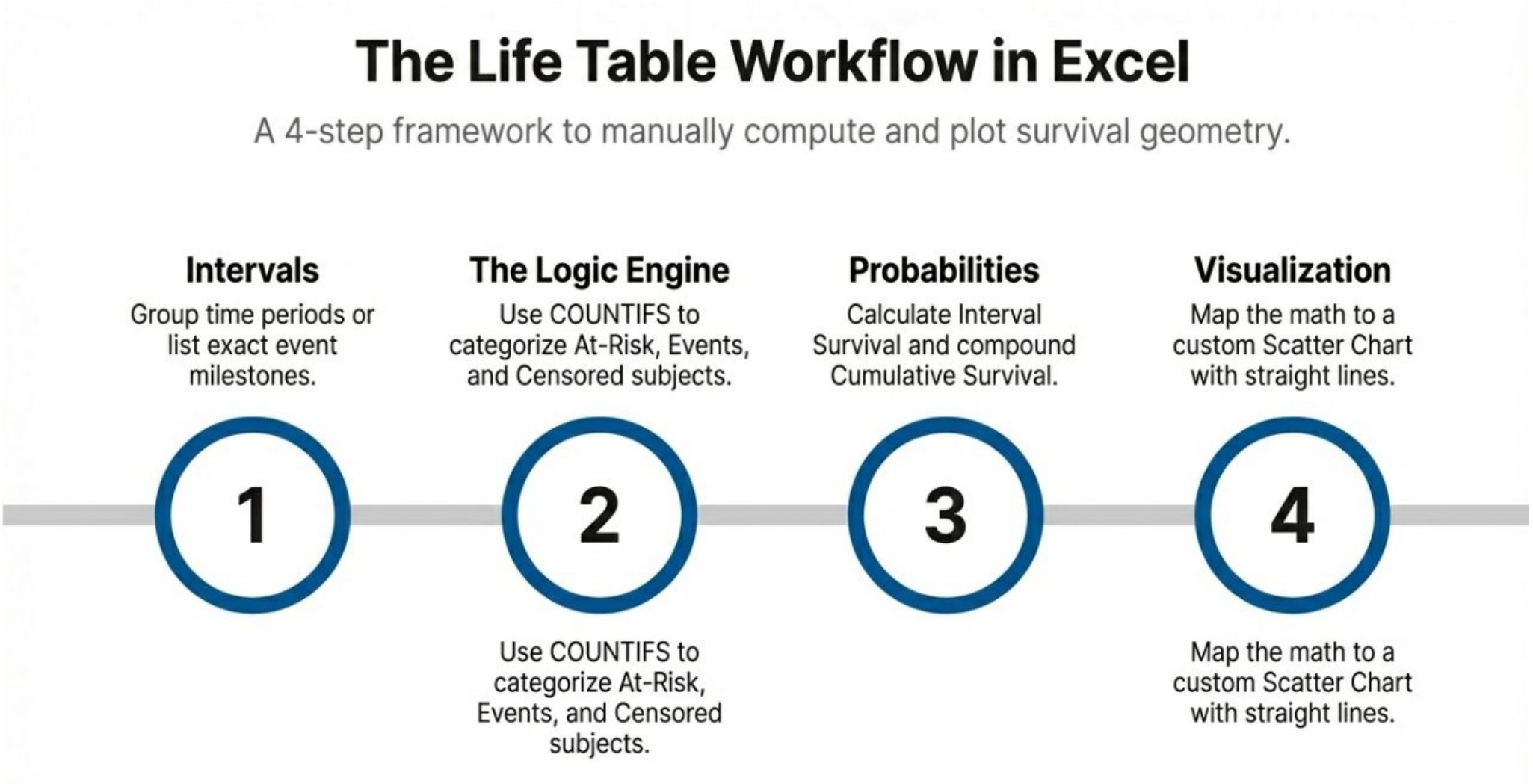
↓

**Step 2:** Identify highly influential covariates (e.g., Treatment vs. Placebo).

↓

**Step 3: Decision Gate:** Determine if separate, stratified KM curves are required for different sub-groups.

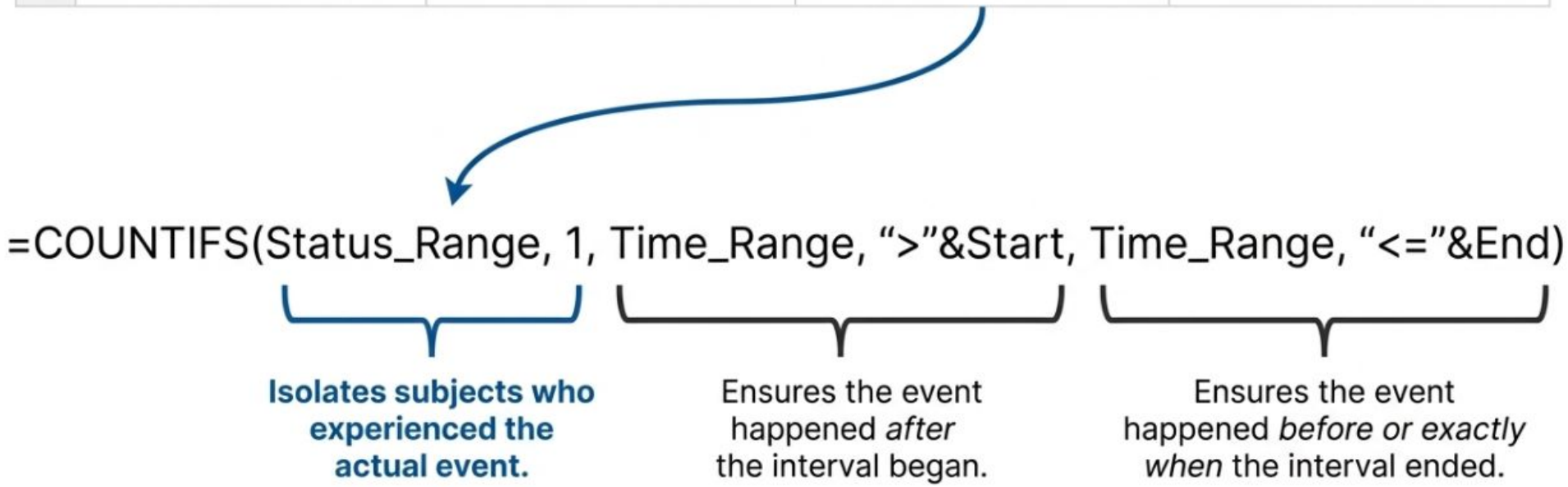
# Survival analysis



# Survival analysis

## Step 1 & 2: Intervals and The Logic Engine

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Time Interval | At Risk | Events (Status=1) | Censored (Status=0) |
| 1 | 0-6 Months | 500 | 42 | 12 |
| 2 | 6-12 Months | 446 | 38 | 15 |



# Survival analysis

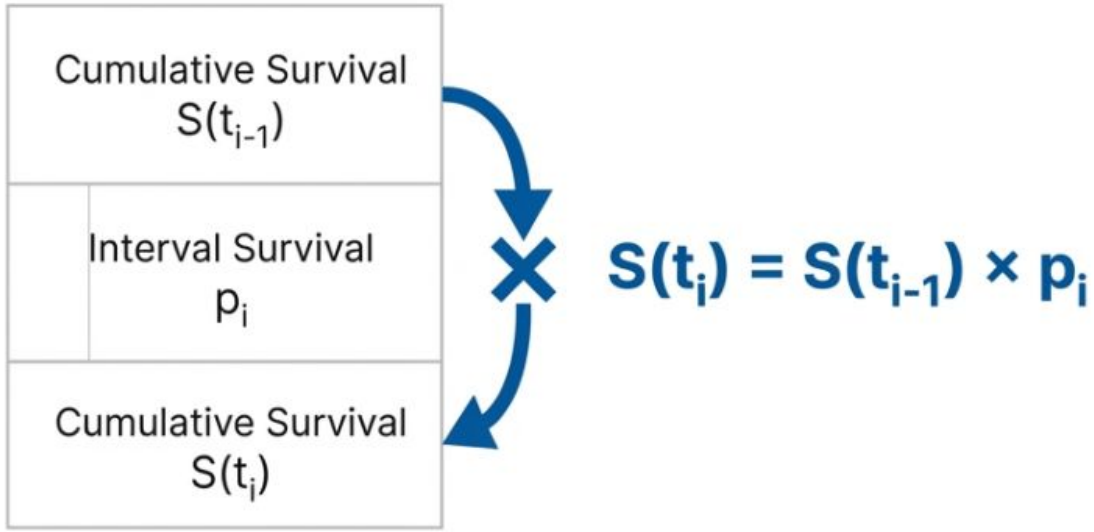
## Step 3: Calculating Probabilities

### Step A: Interval Survival ($p_i$)

$$p_i = 1 - \frac{\text{Events}}{\text{At\_risk}}$$

The probability of surviving *only* this specific interval.

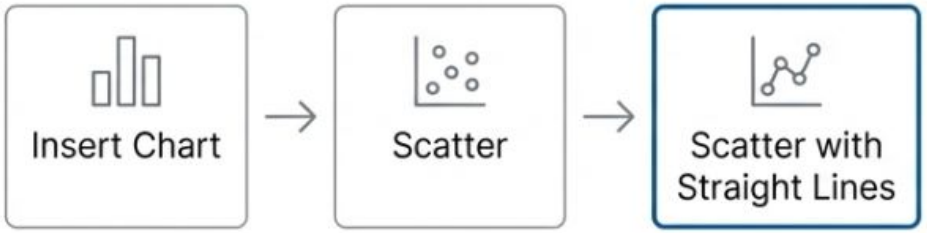
### Step B: Cumulative Survival ($S(t_i)$)




The continuous product of all interval factors up to time t.

# Survival analysis

## Step 4: Visualizing the Curve in Excel




**Axis Data Mapping Protocol**

**X-Axis** (Horizontal) → Map to the '**Time**' column.

**Y-Axis** (Vertical) → Map to the '**Cumulative Survival**' column.

**Pro-Tip:** To create the true 'Step' effect mathematically, ensure your life table data includes rows for the exact moments just before an event occurs.

# Survival analysis

## The Analyst's Master Checklist

| Objective | Excel Execution Methodology |
|---|---|
| Cohort Description | Generate PivotTables for categorical variables. |
| Event & Censored Counting | Deploy =COUNTIFS() with interval bounding logic. |
| Cumulative Probability | Calculate consecutive product of interval factors: $S(t_{i-1}) \times p_i$ |
| KM Curve Generation | Plot via Scatter with Straight Lines (X = Time, Y = Cumulative Survival). |

Survival Analysis Workbook Complete. Reference this matrix for standard execution.

# Thank you!