

UNIT I


Context of data visualization

Data visualization is the representation of data through use of common graphics, such as charts, plots, infographics and even animations. These visual displays of information communicate complex data relationships and data-driven insights in a way that is easy to understand.

Data visualization can be utilized for a variety of purposes, and it's important to note that it is not only reserved for use by data teams. Management also leverages it to convey organizational structure and hierarchy while data analysts and data scientists use it to discover and explain patterns and trends.

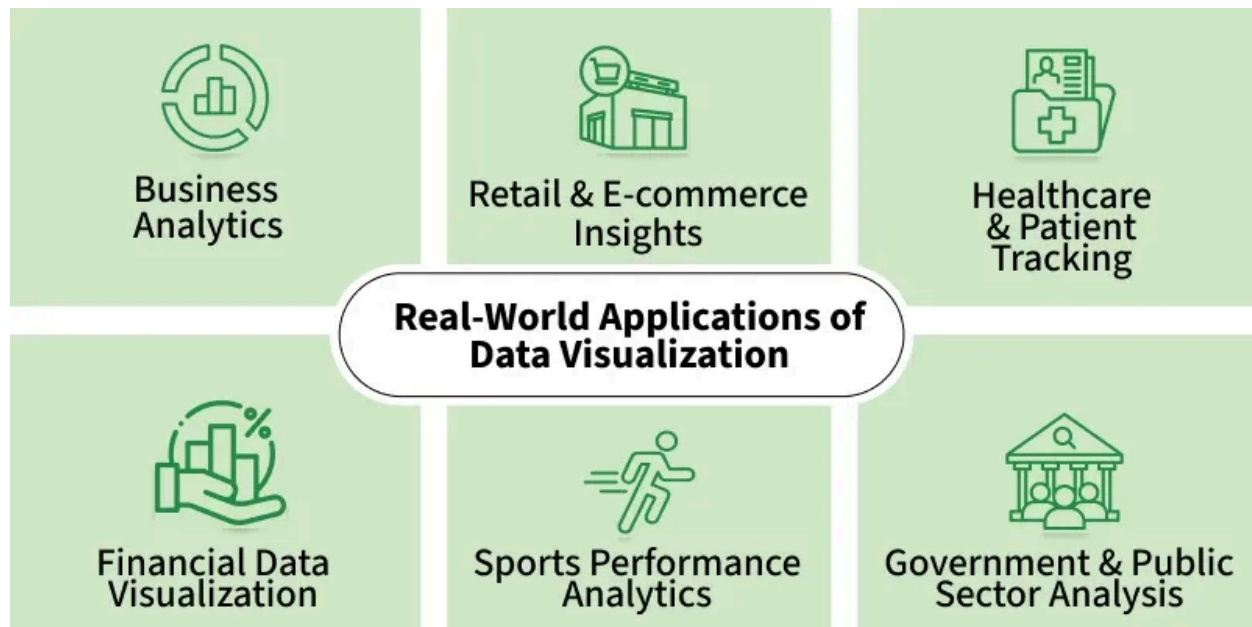
What is Data Visualization?

- Turning data into visual formats like charts, graphs, and maps
- Makes complex data easy to understand at a glance
- Helps identify trends, patterns, and insights



The illustration shows a person in a green sweater and white pants pointing at a large pie chart. A small inset shows a line graph with a wavy line, and another inset shows a bar chart with five bars of varying heights. The person is standing next to the pie chart, which is divided into several segments, including a large black one.

Benefits of Data Visualization



Common Types of Data Visualization

There are various types of visualizations where each has a unique purpose in data representation. Here are the most common types:

1. **Charts and Graphs:** They are used to visualize data, with charts comparing data points across categories or showing trends over time and graphs analyzing relationships between variables to identify correlations, trends and outliers. Examples: Bar Charts, Line Charts, Pie Charts, Scatter Plots, Histograms, Box Plots.
2. **Maps:** They are used to display geographical data which provides spatial context to trends and patterns. Examples: Geographic Maps, Heat Maps
3. **Dashboards:** They combine multiple visualizations into a single interface which provides real-time insights and interactive features for users to explore data.

Importance of Data Visualization

Data visualization is essential for understanding and communicating information effectively. Here are some key reasons why it's important:

1. **Simplifies Complex Data:** It turns large and complicated data into visual formats like charts and graphs, making the information easier to understand.
2. **Reveals Patterns and Trends:** It helps identify trends, relationships and patterns that are not easily seen in raw data or tables.
3. **Saves Time:** Visuals allow quicker interpretation of data, helping users spot key information at a glance instead of manually scanning through numbers.
4. **Improves Communication:** It makes it easier to explain data insights to others, especially those who may not be familiar with the technical details.

5. **Tells a Clear Story:** Data visuals guide the audience through the information step-by-step, making it easier to reach conclusions and make informed decisions.

Real-World Use Cases for Data Visualization

Data visualization is used across various industries to improve decision-making and drive results. Here are a few examples:

1. **Business Analytics:** Used to monitor company performance, track KPIs and make data-driven decisions by visualizing trends, sales and customer metrics.
2. **Healthcare:** Helps in analyzing patient records, tracking disease outbreaks and managing hospital operations through easy-to-read charts and dashboards.
3. **Sports:** Used to visualize player statistics, team performance and match outcomes, helping coaches and analysts improve strategies and training plans.
4. **Retail and E-commerce:** Enables tracking of sales, customer preferences and inventory levels, helping businesses adjust stock and marketing efforts effectively.

Challenges in Data Visualization

1. **Data Quality:** Accuracy of visualizations depends on the quality of the data. If the data is inaccurate or incomplete, the insights from the visualization will be misleading.
2. **Over-Simplification:** Simplifying data too much can lead to important details being lost like using a pie chart that oversimplifies complex relationships between categories.
3. **Choosing the Right Visualization:** Using the wrong type of visualization can distort the message. For example, a pie chart might not work well with many categories which leads to confusion.
4. **Overload of Information:** Too much information in a visualization can overwhelm viewers. It's important to focus on key data points and avoid clutter.

What are the key components of data visualization

In today's world, a huge amount of data is generated every day and it is very important to visualize the data to know its pattern to make important business decisions. At its core, effective *data visualization* relies on several key components, each playing a crucial role in conveying information accurately and efficiently. These components encompass aspects ranging from the choice of visual representation to the use of colour, interactivity, and storytelling techniques.

Understanding these components is essential for creating compelling and informative visualizations that facilitate data-driven decision-making across various domains. In this article, we will learn about *What are the key components of data visualization?*

Main Components of Data Visualization

1. **Data:** First of all we need lots of data. Data can be of any type like numerical data, text data or geospatial data.
2. **Visual Elements:** For instance, Graphics, charts, Overlays, diagrams, figures, maps, tables and other types of data presentations and resumes that comprise infographics.
3. **Visualization Techniques:** This also includes aspects such as transforming and scaling data, and in some cases indeed selecting the right type of visualization to use.
4. **Interactivity:** Influential aspects of visualization which could include; The blinking bubbles whenever the cursor is over a particular part of the graph,

- zooming up or down the size of the graph, or even rotating the graph and/or options to have the different groups or categories on or off among others.
5. **Color Palette and Design:** New choices on color utilization, font selection, position of elements in unison with the style of design to enhance the usability, visibility and the aesthetic appeal of visualization.
 6. **Context and Annotations:** The additional roles of Titles and Subtitles in the Visualizations and some of the other labels for captions, annotations and legends for further understanding of the analysis.
 7. **Tools and Platforms:** Software that falls under "Other Tools and Applications while developing visualization" Some of the prominent ones are Tableau, Power BI and more programming languages and Libraries are D3.js, Matplotlib and more.
 8. **Data Preparation:** Data preparation which involves cleaning of data, that is, data cleaning processing, data selection, data reshaping and data condensation that gets data in the right form so that it can be analyzed and visualized.
 9. **Data Exploration:** Simple methods for analyzing the data for searching the pattern, trend, noise, similarity and correlation other than the techniques.
 10. **Dashboarding:** An aggregate display of the two screens to give an overall view of the data collected as well as endlessly monitoring the indicators by having the two tabbed views on the same screen.

1. Purpose: Exploratory vs. Explanatory

The purpose defines who the visualization is for and what they are supposed to do with it. Most projects fall into one of two categories:

Factor	Exploratory Visualization	Explanatory Visualization
Audience	The Analyst (Internal)	The Decision Maker / Public (External)
Goal	To find the story within the data.	To tell a specific story found in the data.
Interactivity	High (filters, drill-downs).	Low to Moderate (guided paths).
Visual Style	Minimalist, often "aw" or clinical.	Highly polished with clear annotations.

Example: An **Exploratory** dashboard allows a sales manager to filter by region, date, and product to find where things are going wrong. An **Explanatory** slide shows one specific chart proving that "Sales in the North plummeted due to a 20% price hike."

2. Visualization Function: The "What"

The function refers to the specific relationship or pattern you want to reveal. In data design, "**form follows function**"—you should never pick a chart because it looks "cool," but because it performs a specific task.

Common functions include:

- **Comparison:** Showing how values differ (e.g., Bar Charts, Bullet Graphs).
- **Distribution:** Showing how data is spread out (e.g., Histograms, Box Plots).
- **Composition:** Showing how a total is broken down into parts (e.g., Treemaps, Stacked Bars).

-
- **Relationship:** Showing how variables correlate (e.g., Scatter Plots, Bubble Charts).
 - **Trend over Time:** Showing how data changes (e.g., Line Charts, Area Charts).
-

3. Visualization Tone: The "Feel"

The tone is the "personality" of your visualization. It dictates the emotional response and the level of authority the visual carries. Tone is usually established through **typography, color theory, and layout**.

- **Analytical/Assertive:** Minimal color, clean sans-serif fonts, and high data density. Used in scientific papers or executive financial reports.
- **Journalistic/Informative:** High use of annotations, "human" photography or icons, and accessible language. Used in news media (like *The New York Times* or *The Economist*).
- **Artistic/Experimental:** Uses non-traditional forms (like data art) to evoke wonder or curiosity. Often used in museum installations or digital storytelling.
- **Urgent/Cautionary:** Uses high-contrast colors (reds/oranges) and aggressive scaling to signal a crisis, such as a cybersecurity breach or a pandemic surge.

Data Representation vs. Data Presentation

In the context of computer science and data analytics, it is important to distinguish between how data is stored (Representation) and how it is communicated (Presentation).

1. Data Representation

Definition: This refers to the method used to structure, store, and organize data internally within the computer's memory or a file. It is not necessarily meant for human consumption but for efficient processing.

In your Python Code: Data representation occurs when we define the lists (arrays).

- **Categorical Data:** `subjects = ['Math', 'Science', ...]`
- **Numerical Data:** `marks = [85, 92, ...]`

We represent the real-world concept of a "Student's Report Card" using two parallel lists in Python.

2. Data Presentation

Definition: This refers to the method of summarizing and visualizing data to make it understandable and actionable for humans. Good presentation involves choosing the right chart type, colors, labels, and grids.

In your Python Code: Data presentation occurs when we use `matplotlib` to render the visuals.

- **Bar Chart:** Chosen to present a *comparison* between discrete subjects.
- **Line Chart:** Chosen to present a *trend* over time.
- **Enhancements:** Adding titles, axis labels (`plt.xlabel`), and legends are part of the presentation layer, turning raw numbers into meaningful insights.

Summary Table

Concept	Purpose	Python Example

Representation	Internal storage & Structure	<code>sales = [12000, 15000, ...]</code>
Presentation	External communication & Visualization	<code>plt.plot(months, sales)</code>

A 7-Step Data Visualization Framework for Stronger Storytelling

It's far easier for humans to remember images than facts and figures. How many times have your eyes glazed over when looking at a spreadsheet full of numbers? Data visualization helps to solve this problem by presenting the story that the data tells in a pictorial or graphical format. When people see information presented visually, they can more easily digest it, spot patterns and anomalies, and better grasp the insight it reveals. In this article, we're sharing a reliable data visualization framework that you can use to build strong visualizations for more powerful communication.

Step-by-Step Data Visualization Framework

Creating effective visualizations is much easier when you have a step-by-step method for developing them. Following these seven steps will help you improve your data storytelling.

1. Start with a clear goal

Begin by identifying the business question your data analysis answers and decide what you want to achieve with your visualization. What do you want to communicate? What do you want your audience to understand? For example, if you're creating a

visualization for an analysis of last quarter's sales data, your goal might be to reveal which product SKUs are generating the most revenue in each territory.

2. Understand the data

Without a solid framework and understanding of your data, it's easy to make incorrect assumptions that lead to inaccurate conclusions. And these mistakes can lead to costly consequences when decision-makers act on faulty insights. Be sure you understand your data's variables and what each represents, as well as their significance. What variables will help answer the business question? Do you need to bring in any additional data sets to provide a fuller, more accurate picture?

3. Consider your audience

The most effective visualizations are highly focused and fit within a framework. Knowing what to leave out is just as important as knowing what to include. Understanding your audience will help you decide what to focus on. Find out your audience's level of awareness and experience with your subject matter. What do they already know? What information do they need? An audience of subject matter experts will require a very different visualization than an audience new to the subject.

4. Identify which type of visualization is best suited to the data

Different visualization techniques lend themselves to different types of data analysis. For example, charts and graphs are best used for univariate data and descriptive analytics, while diagrams are ideal for demonstrating the complex relationships between hierarchical or multidimensional data, and maps are great for visualizing geographical data. Selecting the right type of visualization that falls within a framework is crucial for effective communication. (A modern data visualization solution like Sigma can help make this important decision for you.)

5. Create your visualization

Once you've decided on the type of visualization your data requires, it's time to design it. Choose the simplest options that best communicate the data's message. Consider strategic use of color to help create mental associations. For example, you might use common color conventions such as orange for safety. Contrasting colors will demonstrate comparison/contrast. You can also use color simply to make important information stand out.

Another thing to consider when creating your visualization is to make it interactive. Interactive visualizations are especially powerful since they enable the user to engage more fully with the data. Allow viewers to manipulate your visualization, highlight key sections, remove what they don't need, and keep what they do. (Sigma makes it easy to keep viewers engaged with interactive dashboards that allow them to freely explore live, underlying data to answer additional questions.)

6. Gather feedback

Testing and gathering feedback will help you make your visualization more effective and by extension your data visualization framework. We all have blind spots, especially when the data is familiar or when we're working in our own area of expertise. Show your visualization to a segment of your target audience and ask for feedback. Is anything unclear or confusing? Could anything be improved to make the message easier to understand?

7. Iterate

Based on the feedback you receive, consider making changes before releasing your visualization to a broader audience. While you don't need to incorporate every suggestion, feedback will often reveal blind spots or opportunities to significantly improve visualization for a specific audience.