UNIT - I

Introduction

In our increasingly data-driven world, it's more important than ever to have accessible ways to view and understand data. After all, the demand for data skills in employees is steadily increasing each year. Employees and business owners at every level need to have an understanding of data and of its impact.

That's where data visualization comes in handy. With the goal of making data more accessible and understandable, data visualization in the form of dashboards is the go-to tool for many businesses to analyze and share information.

1.1 DATA VISUALIZATION

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. Additionally, it provides an excellent way for employees or business owners to present data to non-technical audiences without confusion.

Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. The main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets. The term is often used interchangeably with others, including information graphics, information visualization and statistical graphics.

Data visualization is one of the steps of the data science process, which states that after data has been collected, processed and modelled, it must be visualized for conclusions to be made.

Data visualization is also an element of the broader data presentation architecture (DPA) discipline, which aims to identify, locate, manipulate, format and deliver data in the most efficient way possible.

Data visualization is important for almost every career. It can be used by teachers to display

student test results, by computer scientists exploring advancements in artificial intelligence (AI) or by executives looking to share information with stakeholders. It also plays an important role in big data projects. As businesses accumulated massive collections of data during the early years of the big data trend, they needed a way to get an overview of their data quickly and easily. Visualization tools were a natural fit.

Visualization is central to advanced analytics for similar reasons. When a data scientist is writing advanced predictive analytics or machine learning (ML) algorithms, it becomes important to visualize the outputs to monitor results and ensure that models are performing as intended. This is because visualizations of complex algorithms are generally easier to interpret than numerical outputs.

Unit - I

Introduction to Data Visualization, Key factors of Data Visualization, Importance of Data Visualization in Business Intelligence, Data Visualization tools and types of data, Data Abstraction: data types, dataset types, Attribute types, Task Abstraction, Four Levels for Validation.***Why is data visualization important?***

Data visualization provides a quick and effective way to communicate information in a universal manner using visual information. The practice can also help businesses identify which factors affect customer behaviour; pinpoint areas that need to be improved or need more attention; make data more memorable for stakeholders; understand when and where to place specific products; and predict sales volumes.

Other benefits of data visualization include the following:

✓ the ability to absorb information quickly, improve insights and make faster decisions;

✓ an increased understanding of the next steps that must be taken to improve the organization;

✓ an improved ability to maintain the audience's interest with information they can understand;

✓ an easy distribution of information that increases the opportunity to share insights with

everyone involved;

✓ eliminate the need for data scientists since data is more accessible and understandable; and

✓ an increased ability to act on findings quickly and, therefore, achieve success with greater speed and less mistakes.

1.2 KEYS FACTORS 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. Data visualization is a crucial part of data analytics that helps you visualize your data and uncover significant trends and patterns

2Understanding 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 a 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.

31.3 IMPORTANCE OF DATA VISUALIZATION IN BUSINESS INTELLIGENCE

Data visualization is technique for businesses. It helps them understand their data better, make smarter decisions, and stay ahead of the competition. It basically turns boring numbers into easy-to-understand pictures or graphs, helping businesses see what's going on and what they need to do next.

1. Simplifies Data for Better Understanding: Data visualization makes complex and raw data easy to understand by displaying it in the form of graphs and charts. This helps us see patterns and important information more clearly.

2. Help Us Make Better Decisions: When we can see our data in pictorial form or graphs and chart, it's easier to make smart decisions quickly.

3. Helps In Visualizing Data Easily: Even the people who are not experts in data can understand data easily with the help of visualization. Non-technical people can also understand the data easily with the help of visualization tools, in the form of graphs and charts.

4. Makes Data More Interesting: Visualizations make data more interesting and fun to look at, which encourages more people to use data to help their work.

5. Helps in Tracking Progress: Visual dashboards keeps the track of our progress. They help us see if we're reaching our goals by giving us a clear picture of our progress.

1. 4 DATA VISUALIZATION TOOLS AND TYPES OF DATA

Data visualization tools are cloud-based applications that help you to represent raw data in easy-to-understand graphical formats. You can use these programs to produce customizable bar charts, pie charts, column charts, and more

There are many tools that are used for data visualization. Some of the tools are discussed below

1) Power BI

Power BI is a Business Intelligence and Data Visualization tool which helps you to convert data from various data sources into interactive dashboards and reports. It provides multiple software connectors and services.

• Power BI is a business analytics tool by Microsoft that provides interactive visualizations and business intelligence capabilities.

• With the help of power BI we can structure data and make business decisions out of those insights.

• It allows users to connect to a wide range of data sources, create interactive reports and dashboards, and share them with others.

42) Tableau

Tableau is a robust tool for visualizing data in a better way. You can connect any database to create understandable visuals. It is one of the best visualization tools that enables you to share visualization with other people.

• Tableau is one of the most popular data visualization tools.

• It is capable of learning the store's business patterns amd running queries against the data to help visualize the flaws and resolve them quickly.

• It allows users to create interactive and shareable dashboards, reports, and charts.

• Tableau supports various data sources and offers a user-friendly interface for creating visualizations.

3) Qlik

Qlik is a data visualization software which is used for converting raw data into knowledge. This software acts like a human brain which works on "association" and can go into any direction to search the answers.

• **QlikView and Qlik Sense** are data visualization and business intelligence tools developed by Qlik.

• They allow users to create interactive visualizations, dashboards, and reports using data from multiple sources.

• QlikView helps us to understand complex trends, patterns and convert it itno actionable insights.

• Qlik Sense is more modern and user-friendly compared to QlikView.

4) Google Data Studio

• Google Data Studio is a free data visualization tool that allows users to create interactive dashboards and reports.

• It is a dashboard and reporting tool that is easy to use, customize and share.

• It integrates seamlessly with other Google products such as Google Analytics, Google Sheets, and Google BigQuery.

Types of data

The data is classified into majorly four categories:

1. Nominal data

2. Ordinal data

3. Discrete data

4. Continuous data

Further, we can classify these data as follows:

5 *Qualitative or Categorical Data*

Qualitative data, also known as the categorical data, describes the data that fits into the categories. Qualitative data are not numerical. The categorical information involves categorical variables that describe the features such as a person's gender, home town etc. Categorical measures are defined in terms of natural language specifications, but not in terms of numbers. Sometimes categorical data can hold numerical values (quantitative value), but those values do not have a mathematical sense. Examples of the categorical data are birthdate, favourite sport, school postcode. Here, the birthdate and school postcode hold the quantitative value, but it does not give numerical meaning.

Nominal Data

Nominal data is one of the types of qualitative information which helps to label the variables without providing the numerical value. Nominal data is also called the nominal scale. It cannot be ordered and measured. But sometimes, the data can be qualitative and quantitative. Examples of nominal data are letters, symbols, words, gender etc.

The nominal data are examined using the grouping method. In this method, the data are grouped into categories, and then the frequency or the percentage of the data can be calculated. These data are visually represented using the pie charts.

Ordinal Data

Ordinal data/variable is a type of data that follows a natural order. The significant feature of the nominal data is that the difference between the data values is not determined. This variable is mostly found in surveys, finance, economics, questionnaires, and so on. The ordinal data is commonly represented using a bar chart. These data are investigated and interpreted through many visualisation tools. The information may be expressed using tables in which each row in the table shows a distinct category.

Quantitative or Numerical Data

6Quantitative data is also known as numerical data which represents the numerical value (i.e., how much, how often, how many). Numerical data gives information about the quantities of a specific thing. Some examples of numerical data are height, length, size, weight, and so on. Quantitative data can be classified into two different types based on the data sets. The two different classifications of numerical data are discrete data and continuous data.

Discrete Data

Discrete data can take only discrete values. Discrete information contains only a finite number of possible values. Those values cannot be subdivided meaningfully. Here, things can be counted in whole numbers.

Example: Number of students in the class

Continuous Data

Continuous data is data that can be calculated. It has an infinite number of probable values that can be selected within a given specific range.

Example: Temperature range

71.5 DATA ABSTRACTION

Data abstraction is the process of structuring, categorizing, and interpreting raw data into meaningful forms that can be effectively visualized and analyzed.

Key Aspects of Data Abstraction includes

1. Data Types

• **Item**: A single, discrete entity (e.g., a row in a table).

• **Attribute**: A measurable property or variable (e.g., income, temperature).

• **Link**: A relationship between items (e.g., social connections).

• **Position**: Spatial data defined by coordinates.

• **Grid**: Sampling strategy for continuous data in a structured layout.

2. Dataset Types

• **Tables**: 2D representation; rows = items, columns = attributes.

- **Networks**: Nodes (items) connected via links (relationships).

- **Fields**: Continuous domain data sampled at discrete points.

- **Geometry**: Describes shape, structure, and spatial layout (points, surfaces, volumes).

- **Other**: Clusters

A Brief History of Data Visualization

Data visualization is the graphical representation of data to communicate information clearly and effectively. Its history dates back several centuries.

- **17th century:** Early maps and diagrams were used to represent geographic and scientific data.

- **18th century:** William Playfair introduced **bar charts, line graphs, and pie charts**, which are still widely used today.

- **19th century:** Visualization became important in statistics and public health.
  Example: **John Snow's cholera map (1854)** showed how disease spread through contaminated water.

- **20th century:** Computers enabled faster data processing and graphical representation.

- **21st century:** Modern tools like **Tableau, Power BI, Python, and R** allow interactive and real-time visualization of large datasets.

Data visualization has evolved from simple hand-drawn charts to advanced digital dashboards.

Good Graphics

Good graphics present information **clearly, accurately, and efficiently**. A well-designed visualization helps users understand data quickly.
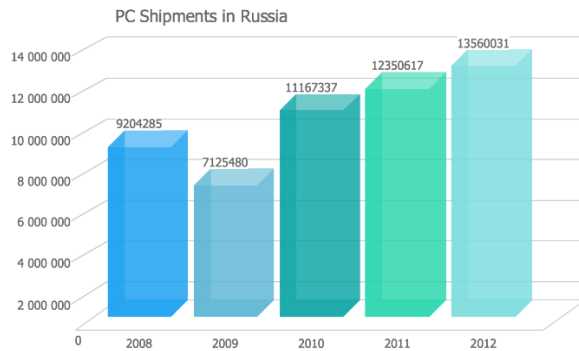
Characteristics of good graphics:

- **Clarity:** Easy to read and understand

- **Accuracy:** Represents data without distortion

- **Simplicity:** Avoids unnecessary decoration

- **Proper labeling:** Titles, legends, and axes must be clear

- **Consistency:** Uses uniform colors, scales, and symbols

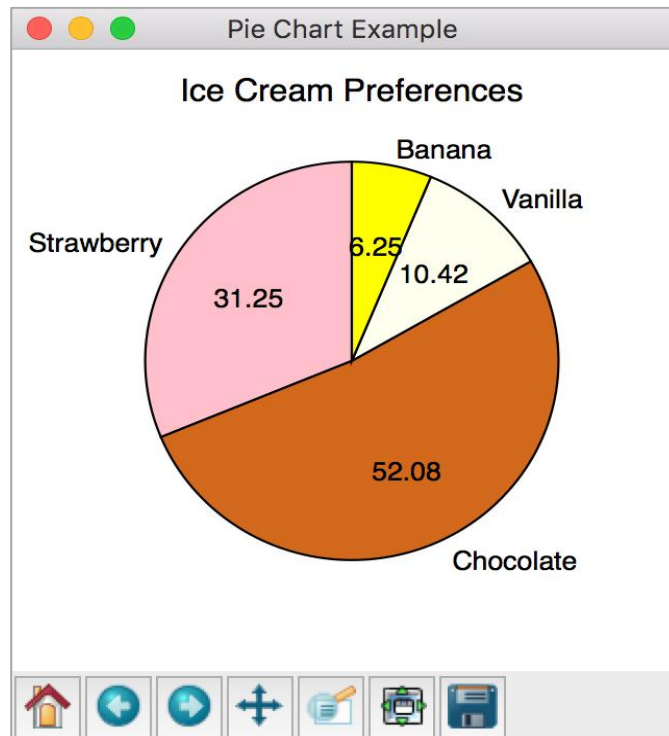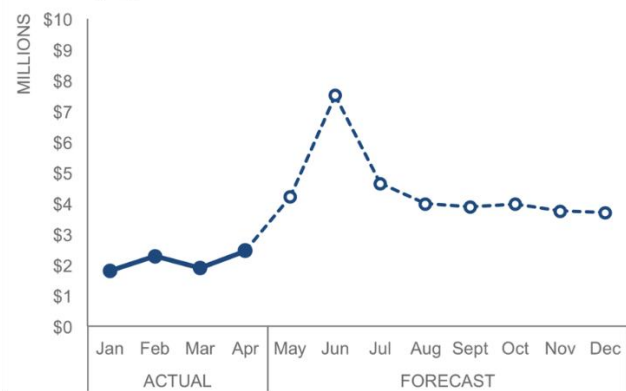- **Relevance:** Focuses only on important data

Edward Tufte emphasized that good graphics should **maximize data understanding while minimizing unnecessary design elements**.

Importance of Good Graphics

- Helps in quick decision-making
- Improves communication of complex data
- Makes reports more understandable
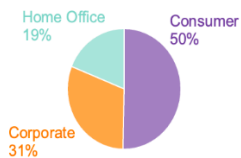- Reduces misinterpretation of data

PC Shipments in Russia
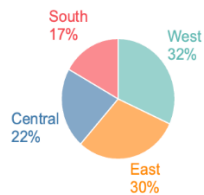
14 000 000
12 000 000
10 000 000
8 000 000
6 000 000
4 000 000
2 000 000
0

9204285
7125480
11167337
12350617
13560031

2008 2009 2010 2011 2012

Monthly spend

MILLIONS
$10
$9
$8
$7
$6
$5
$4
$3
$2
$1
$0

Jan Feb Mar Apr May Jun Jul Aug Sept Oct Nov Dec
ACTUAL   FORECAST

Pie Chart Example

Ice Cream Preferences

Banana
Vanilla
Strawberry
6.25
10.42
31.25
52.08
Chocolate

✅ **Do**

% of Sales in Segments

Home Office
19%
Consumer
50%
Corporate
31%

❌ **Don't**

% of Sales in Regions

South
17%
West
32%
Central
22%
East
30%

Slices are too similar to each other

Static Graphics

Static graphics are **non-interactive visual representations of data**. They do not change once created.

Examples:

- Bar charts

- Line graphs

- Pie charts

- Printed reports

- Images in books and newspapers

Advantages:

- Easy to create

- Suitable for reports and presentations

- Do not require software interaction

Disadvantages:

- No user interaction

- Cannot explore data dynamically

- Limited ability to display large datasets

Static graphics are commonly used in **academic reports, textbooks, and printed documents**.

Example to Draw in Exam

Draw a **line graph showing monthly rainfall**:

- X-axis → Months

- Y-axis → Rainfall

- Join points with a line

Since it does not change, it is a **static graphic**.

Bar Chart

	Europe	U.S.	Japan
Qtr 4			
Qtr 3			
Qtr 2			
Qtr 1			

Axis: $0 $20,000 $40,000 $60,000

Push-ups

NUMBER OF PUSH-UPS
30
25 25   25
20   20
15   15
10   10
5   5
0   0

Sunday Monday Tuesday Wednesday Thursday Friday Saturday
DAYS

— Push-ups

©Study.com

WEEKLY POLL

WE ASKED:

Should the national park authority sell off land assets like Stickle Tarn?
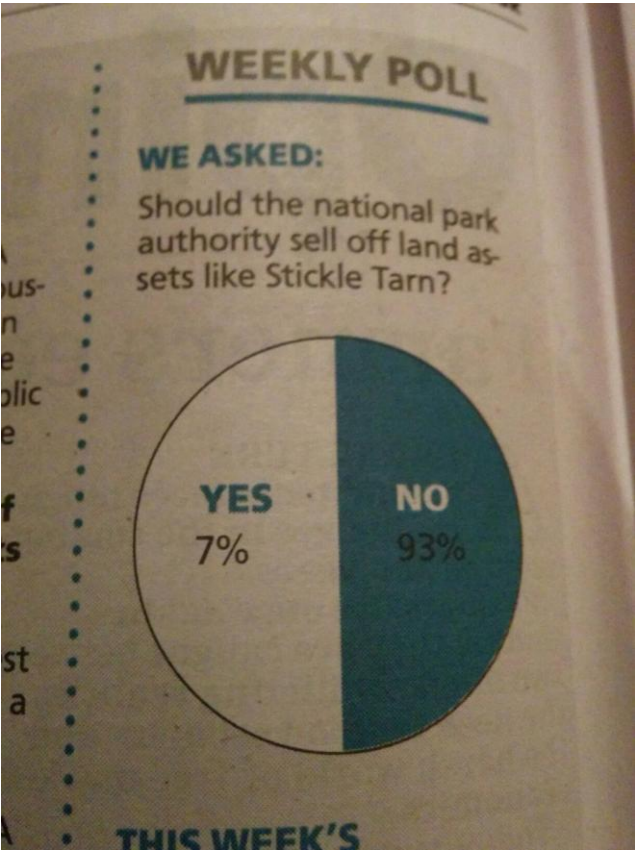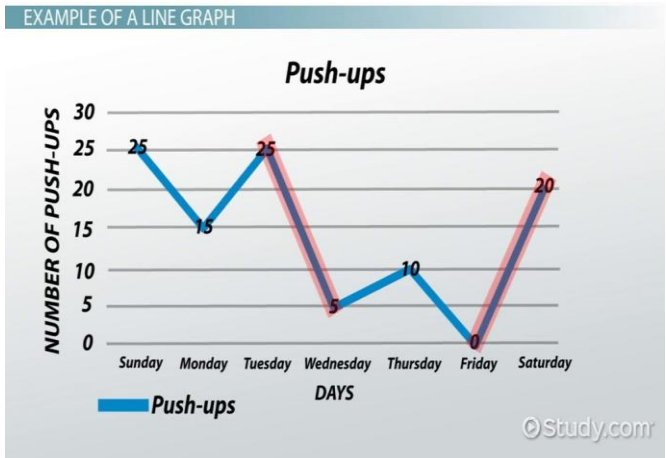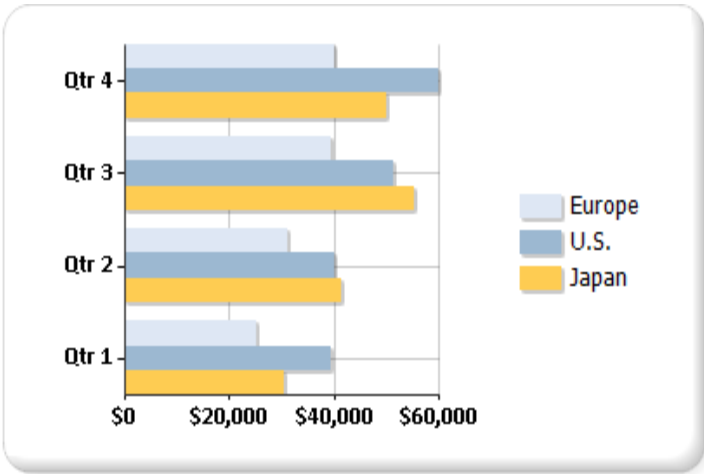
YES 7% NO 93%

THIS WEEK'S

Figure2. A pie chart that displays the number of news items ...

6 6 2 1

- Social news
- Political news
- Sports news
- Exclusive news
- Judicial news

60