BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Introduction to Big Data Analytics

Course Code: 23ST02VAC

Unit-III

DATA VISUALIZATION

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

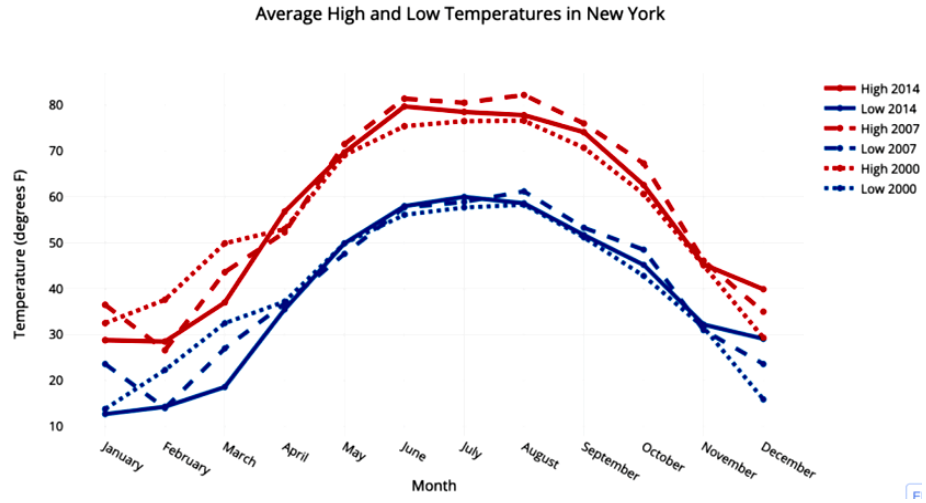
Ms. S. Soundarya
Guest Faculty
Department of Statistics

INTRODUCTION TO BIG DATA ANALYTICS

UNIT –III

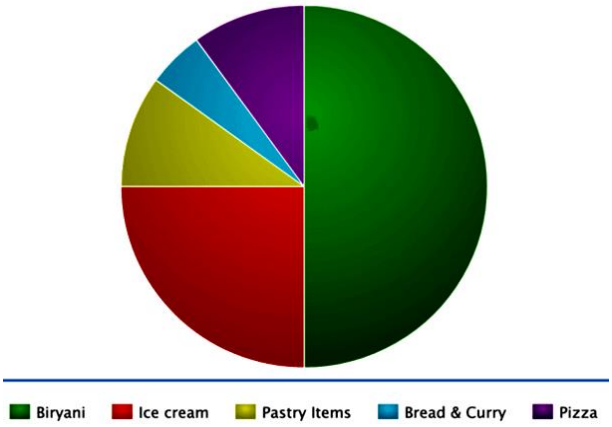
DATA VISUALIZATION

The representation of data or information in graphical means so that the viewer can easily understand the process of trends in the data is called data visualization. Any visual elements are used so that the trends, outliers, and patterns can be understood and studied from the data. Different tools used for data visualization include charts, maps, and graphs, and so on. Tools are designed in such a way to easily understand and grasp the information provided by the data rather than going through the entire process of scanning the excel sheets so that time is saved for the analyst.



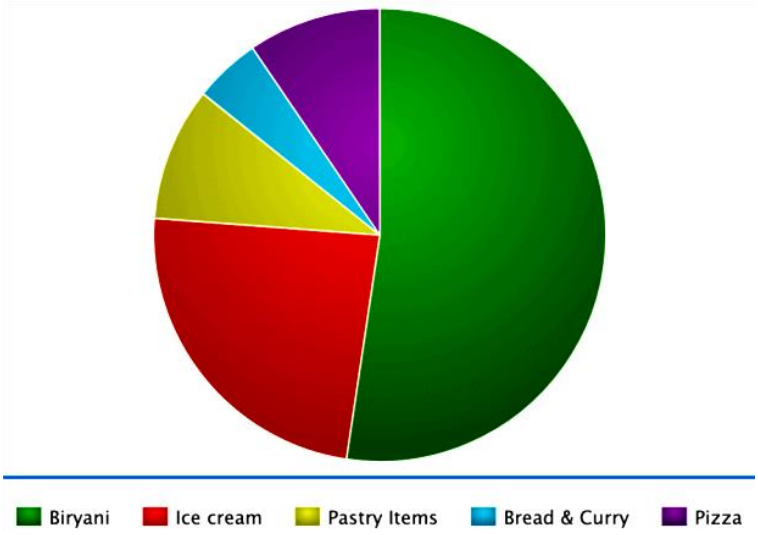
Pie Charts

One of the most ubiquitous charts in use is the food chart or the pie chart.

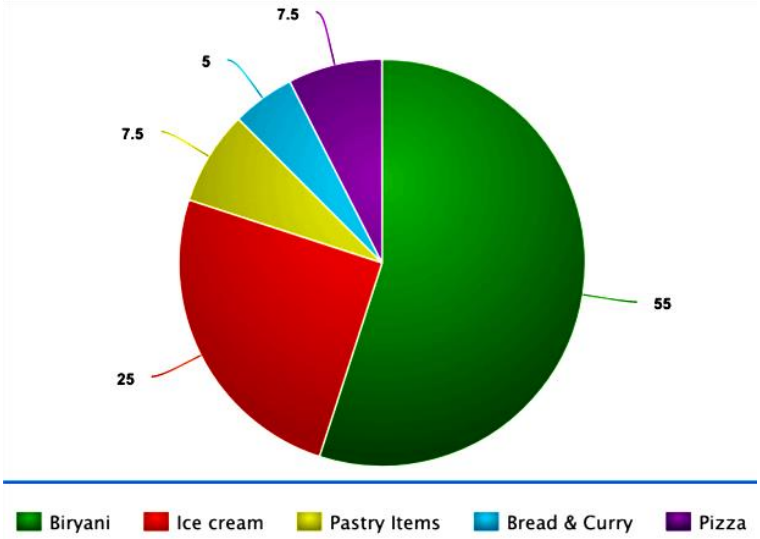


The proportion of Gases in a mixture

Commonly used during elections to show which party has got the lions to share of votes, this chart, although extremely popular, has some glaring disadvantages. Statisticians like Edward Tufte, Leland Wilkinson, and Gerald van Belle have expressed their disinclination towards the use of pie charts. The most conclusive evidence against pie charts has probably been provided by Stephen Few, the founder of Perceptual Edge. In his article, "Save the Pie for the Dessert," published in the Visual Business Intelligence Newsletter, Stephen explains why, to the surprise of his colleagues and students, he feels that pie charts are an ineffective means of visualizing data. Take a look at the pie chart below.

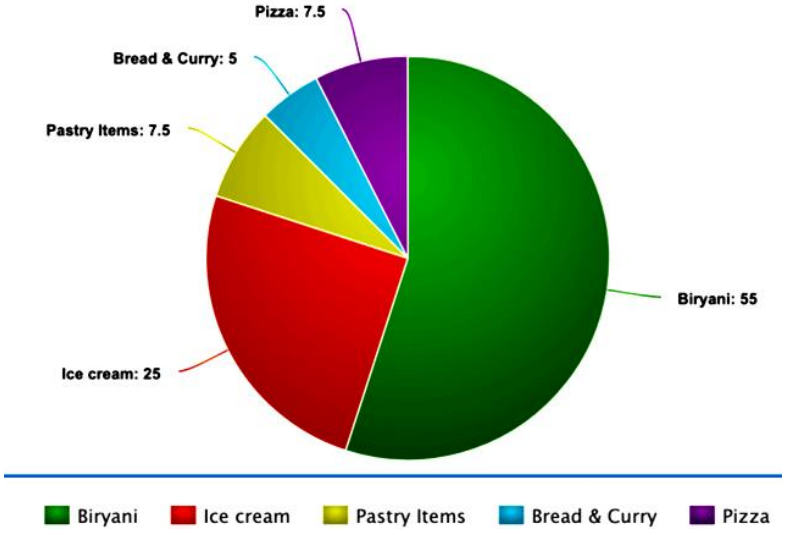


In this graph depicting the proportion of food items ordered after 9 pm, it is easy to estimate the percentage belonging to Ice Cream — 25%. However, after slight modification, it is no longer possible to easily ascertain the percentage of orders going for ice creams. Although the percentage of ice creams is still the same at 25%, it is difficult to understand from the diagram above. In the previous diagram, it was easy to guess the percentage because the shade for ice cream began at the 6 o'clock position and stretched till the 9 o'clock position. As our eyes are trained to see the clocks at 12, 3, 6, and 9 — a pie chart is also easier to interpret if its regions adhere to the 4 clock positions. An avid pie chart supporter would come forward and suggest that this problem could easily be avoided by using labels and marking the percentage against each region.



Pie chart with data labels

But then again, in order to map the value with the item, our eyes have to keep shifting focus from the legend at the bottom of the graph to the labels on the chart. The solution to that would be to attach the item name also, along with the values — to get something similar to this:



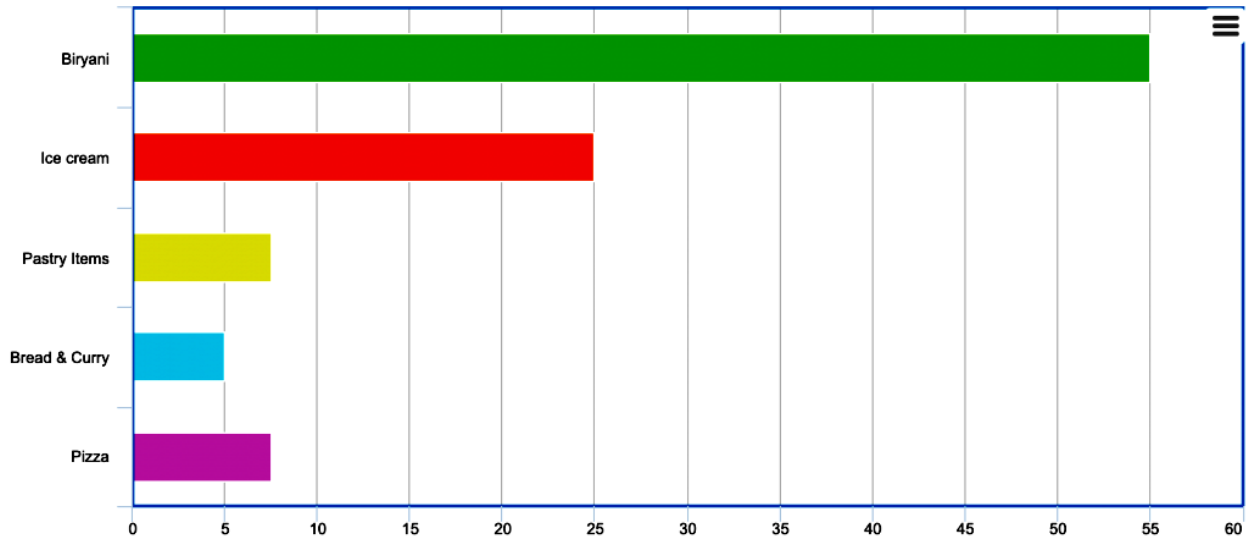
Perfect, you'd say. But then I ask you — Now that we have attached the names of the items along with the values, how is the pie chart above different from the table below:

Food Type	Percentage Of Orders
Biryani	55
Ice cream	25
Pastry Items	7.5
Pizza	7.5
Bread & Curry	5

Clean table with data sorted on Percentage of order. Does this information not look much better than the labels and legends scattered across the pie chart. Visualizations are supposed to add value; they are agents of conclusions; they are supposed to make the data expose its secrets. But in this case, the pie chart does not fulfill its promise.

Enter the Bar Chart of data visualization

The leaner and much more elegant alternative to the pie chart is the bar chart, which is able to express all that the pie says and much more without making a mess of labels and legends.



Bar chart depicting the same — easy to compare, contrast, and conclude This is the same information now presented gently to the observer. Besides showing the individual values, this graph enables us to quickly compare the popularity of each food item and even compare the relative difference in value between consecutive items. Same data would have been available through the table, but our brains are better at processing images than deciphering the text. This is why the use of the pie chart has reduced over time. Although a favorite visual in the media industry, this has been replaced by better and more sophisticated representations. Besides line graphs, pies, and bar charts, there is a myriad of other charts that come in handy. One interesting graph, but not so popular to the uninitiated, is the Sankey Diagram. This diagram is used to express the distribution and flow of matter. The matter in question could be energy, money, or even water.

TOOLS

List of Big Data Analytics Tools

Data Analytics is the process of analysing datasets to draw results, on the basis of information they get. It is popular in commercial industries, scientists and researchers to make a more informed business decision and to verify theories, models and hypothesis. These are the 10 Best Big Data Analytic Tools with their uses and limitations, which can help you to analyse the data. Let's discuss them one by one:

- Tableau Public
- OpenRefine
- KNIME
- RapidMiner
- Google Fusion Tables
- NodeXL
- Wolfram Alpha
- Google Search Operators
- Solver
- Dataiku DSS

- **Tableau Public**

Tableau Public – Big Data Analytics Tools: It is a simple and intuitive tool. As it offers intriguing insights through data visualization. Tableau Public's million-row limit. As it's easy to use fares better than most of the other players in the data analytics market. With Tableau's visuals, you can investigate a hypothesis. Also, explore the data, and cross-check your insights.

Uses of Tableau Public

- You can publish interactive data visualizations to the web for free.
- No programming skills required.

  Visualizations published to Tableau Public can be embedded into blogs. Also, web pages and be shared through email or social media. The shared content can be made available s for downloads. This makes it the best Big Data Analytics tools.

Limitations of Tableau Public

- All data is public and offers very little scope for restricted access
- Data size limitation
- Cannot be connected to R.
- The only way to read is via O Data sources, is Excel or txt.

- **OpenRefine**

OpenRefine – Data Analytic Tools: Formerly known as Google Refine, the data cleaning software. As it helps you clean up data for analysis. It operates on a row of data. Also, have cells under columns, quite similar to relational database tables.

Uses of OpenRefine

- Cleaning messy data
- Transformation of data
- Parsing data from websites

  Adding data to the dataset by fetching it from web services. For instance, OpenRefine could be used for geocoding addresses to geographic coordinates

Limitations of OpenRefine

- Open Refine is unsuitable for large datasets.
- Refine does not work very well with big data

- **KNIME**

KNIME – Data Analysis Tools: KNIME helps you to manipulate, analyze, and model data through visual programming. It is used to integrate various components for data mining and machine learning.

Uses of KNIME

- Don't write blocks of code. Rather, you have to drop and drag connection points between activities.
- This data analysis tool supports programming languages. In fact, analysis tools like these can be extended to run chemistry data, text mining, python, and R.

Limitation of KNIME

- Poor data visualization

- **RapidMiner**

RapidMiner – Data Analytic Tools: RapidMiner provides machine learning procedures. And data mining including data visualization, processing, statistical modeling and predictive analytics. RapidMiner written in Java is fast gaining acceptance as a Big data analytics tool.

Uses of RapidMiner

- It provides an integrated environment for business analytics, predictive analysis.
- Along with commercial and business applications, it is also used for application development.

Limitations of RapidMiner

- RapidMiner has size constraints with respect to the number of rows.
- For RapidMiner, you need more hardware resources than ODM and SAS.

- **Google Fusion Tables:** When comes to data tools, we have a cooler, larger version of Google Spreadsheets. An incredible tool for data analysis, mapping, and large dataset visualization. Also, Google Fusion Tables can be added to business analytics tools list. This is also one of the best Big Data Analytics tools.

  Uses of Google Fusion Tables

  - Visualize bigger table data online.
  - Filter and summarize across hundreds of thousands of rows.
  - Combine tables with other data on the web
  - You can merge two or three tables to generate a single visualization that includes sets of data.
  - You can create a map in minutes!

  Limitations of Google Fusion Tables

  - Only the first 100,000 rows of data in a table are included in query results or mapped.
  - The total size of the data sent in one API call cannot be more than 1MB.

- **NodeXL:** It is a visualization and analysis software of relationships and networks. NodeXL provides exact calculations. It is a free (not the pro one) and open-source network analysis and visualization software. NodeXL is one of the best statistical tools for data analysis. In which includes advanced network metrics. Also, access to social media network data importers, and automation.

  Uses of NodeXL

  This is one of the data analysis tools in Excel that helps in the following areas:

  - Data Import
  - Graph Visualization
  - Graph Analysis
  - Data Representation

  This software integrates into Microsoft Excel 2007, 2010, 2013, and 2016. It opens as a workbook with a variety of worksheets containing the elements of a graph structure. That is like nodes and edges. This software can import various graph formats. Such adjacency matrices, Pajek .net, UCINet .dl, GraphML, and edge lists.

  Limitations of NodeXL

  - You need to use multiple seeding terms for a particular problem.
  - Running the data extractions at slightly different times.

- **Wolfram Alpha:** It is a computational knowledge engine or answering engine founded by Stephen Wolfram.

  Uses of Wolfram Alpha

  - Is an add-on for Apple's Siri
  - Provides detailed responses to technical searches and solves calculus problems.
  - Helps business users with information charts and graphs. And helps in creating topic overviews, commodity information, and high-level pricing history.

  Limitations of Wolfram Alpha

  - Wolfram Alpha can only deal with a publicly known number and facts, not with viewpoints.
  - It limits the computation time for each query.

- **Google Search Operators**

It is a powerful resource which helps you filter Google results. That instantly to get most relevant and useful information.

Uses of Google Search Operators

- Faster filtering of Google search results
- Google's powerful data analysis tool can help discover new information.

- **Solver**

Excel Solver

The Solver Add-in is a Microsoft Office Excel add-in program. Also, it is available when you install Microsoft Excel or Office. It is a linear programming and optimization tool in excel. This allows you to set constraints. It is an advanced optimization tool that helps in quick problem-solving.

Uses of Solver

- the final values found by Solver are a solution to interrelation and decision.
- It uses a variety of methods, from nonlinear optimization. And also linear programming to evolutionary and genetic algorithms, to find solutions.

Limitations of Solver

- Poor scaling is one of the areas where Excel Solver lacks.
- It can affect solution time and quality.
- Solver affects the intrinsic solvability of your model.

- **Dataiku DSS**

This is a collaborative data science software platform. Also, it helps a team build, prototype, explore. Although, it deliver their own data products more efficiently.

Uses of Dataiku DSS

Dataiku DSS– Data analytic tools provide an interactive visual interface. As in this they can build, click, and point or use languages like SQL.

Limitation of Dataiku DSS

- Limited visualization capabilities
- UI hurdles: Reloading of code/datasets
- Inability to easily compile entire code into a single document/notebook
- Still, need to integrate with SPARK. These were the top data analytics tools and this is all on Best Big Data Analytics tools.

Introduction to Types of Data Visualization

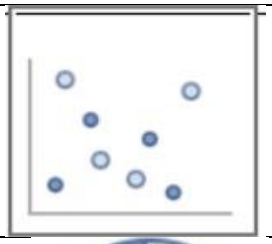
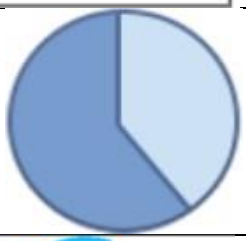
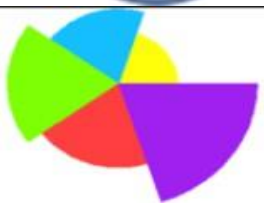
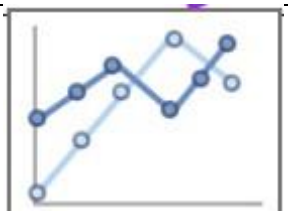
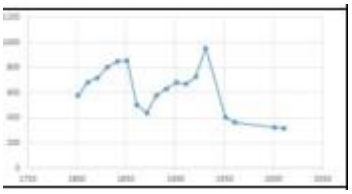
Data Visualization is defined as the pictorial representation of the data to provide the fact- based analysis to decision-makers as text data might not be able to reveal the pattern or trends needed to recognize data; based upon the visualization, it is classified into 6 different types, i.e. Temporal (data is linear and one dimensional), Hierarchical (it visualizes ordered groups within a larger group), Network (involve visualization for the connection of datasets to datasets), Multidimensional (contrast of temporal type), Geospatial(involves geospatial or spatial maps) and Miscellaneous.

Data Visualization

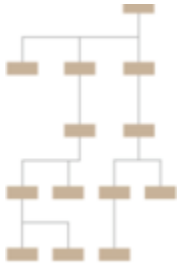
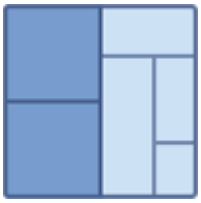
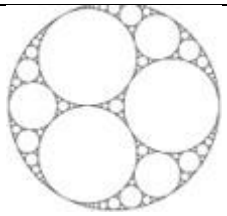
Data visualization is a methodology by which the data in raw format is portrayed to bring out the meaning of that. With the advent of big data, it has become imperative to build a meaningful way of showcasing the data so that the amount of data doesn't become overwhelming. The part of portrayin the data can be used for various purposes, such as finding trends/commonalities/patterns in data, building models for machine learning, or being used for a simple operation like aggregation.

Different Types of Data Visualization

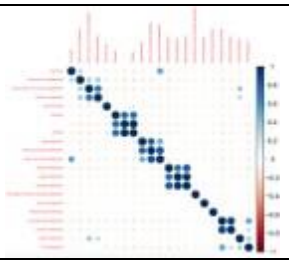
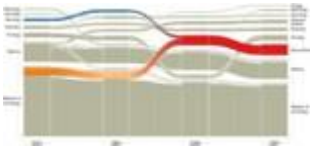
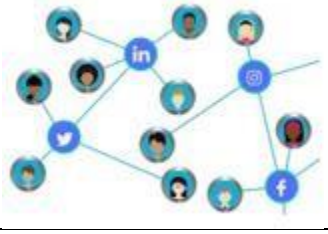
Data visualization is broadly classified into 6 different types. Though the area of data visualization is ever-growing, it won't be a surprise if the number of categories increases.

Temporal: Data for these types of visualization should satisfy both conditions: data represented should be linear and should be one dimensional. These visualization types are represented through lines that might overlap and have a common start and finish data point.		
Scatter Plots	Uses dots to represent a data point. The most common in today's world is machine learning During exploratory data analysis.	
Pie Chart	This type of visualization includes circular graphics where the arc length signifies the magnitude.	
Polar area diagram	Like Pie chart, the Polar area diagram is a circular plot, except the sector angles are equal in length, and the distance of extending from center signifies the magnitude.	
Line graphs	Like the scatter plot, the data is represented by points, except joined by lines to maintain continuity.	
Time lines	In this way, we display a list of data points in chronological order of time.	
Time series sequences	In time series, we represent the magnitude of data in a 2-Dgraphinchronological order of timestamp in data.	

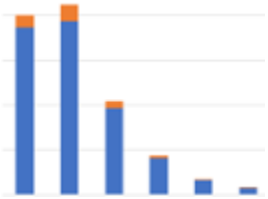
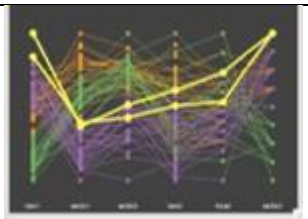
Hierarchical: These types of visualizations portray ordered groups within a larger group. In simple language, the main intuition behind these visualizations is the clusters can be displayed if the flow of the clusters starts from a single point.

Tree Diagram	In a tree diagram, the hierarchical flow is represented in the form of a tree, as the name suggests. Few terminologies for this representation are: – Root Node: Origination point. – Child node: Hasaparent above – Leaf node: Nomorechild node.	
Ring Charts/Sunburst Diagram	The tree representation in the Tree diagram is converted into a radial basis. This type helps in presenting the tree in a concise size. The innermost circle is the root node. And the area of the child node signifies the % of data.	
Tree Map	The tree is represented in the form of rectangles closely packed. The area signifies the quantity contained.	
Circle Packing	Similar to a tree map, it uses circular packing instead of rectangles.	

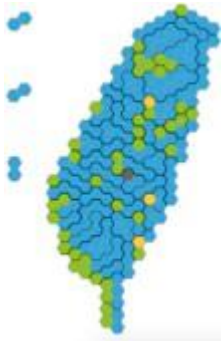
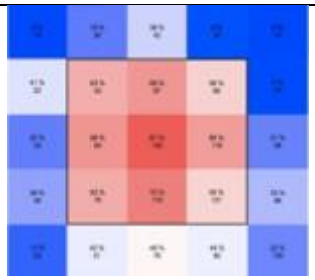
Network: The visualization of these type connects datasets to datasets. These visualizations portray how these datasets relate to one another within a network.

Matrix charts	This type of visualization is widely used to find the connection between different variables within themselves. For example, correlation plot.	
Alluvial diagrams	This is a type of flow diagram in which the changes in the flow of the network are represented over intervals as desired by the user.	
Word cloud	This is typically used for representing text data. The words are closely packed, and the size of the text signifies the frequency of the word.	
node-link diagrams	Here the nodes are represented as dots, and the connection between nodes is presented.	

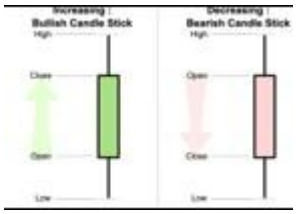
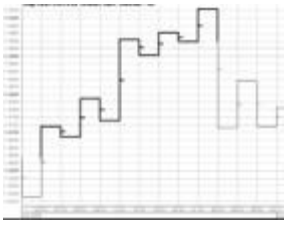
Multidimensional: In contrast to the temporal type of visualization, these types can have multiple dimensions. In this, we can use 2 or more features to create a 3-D visualization through concurrent layers. These will enable the user to present key takeaways by breaking a lot of non-useful data.

Scatter plots	In multi-dimensional data, we select any 2 features and then plot them in a 2-D scatter plot. Bydoingthiswe would have nC2 = n(n-1)/2 graphs.	
Stacked bar graphs	The representation segment bars on top of each other. It canbeeithera100%Stacked Bar graph where the segregation is represented in % or a simple stacked bar graph, which denotes the actual magnitude.	
Parallel Co-ordinate plot	In this representation, a backdrop is drawn, and n parallel lines are drawn(for n-dimensional data).	

Geospatial: These visualizations relates to present real-life physical location by crossing it over with maps (It may be a geospatial or spatial map). The intuition behind these visualizations is to create a holistic view of performance.

Flow map	The movement of information or objects from one location to another is presented where the size of the arrow signifies the amount.	
Choropleth Map	The geospatial map is colored on thebasisofa particular data variable.	
Cartogram	This type of representation uses the thematic variable for mapping. These maps distort reality to present information. This means that on a particular variable, the maps are exaggerated. For example, the image on the left is a spatial map distorted to a bee-hive structure.	
Heat Map	These are very similar to **Choropleth** in the Geospatial genre but can be used in areas apart from geospatial as well.	

Miscellaneous: These visualizations can't be generalized in a particularly large group. So instead of forming smaller groups for the individual type, we group it into miscellaneous. Few examples are below:

Open-High-Low-Close chart	This type of graphs is typically used for stock price representation. The increasing trend is called As **Bullish** and decreasing as **Bearish.**	
Kagi-Chart	Typically the demand-supply of an asset is represented using this chart.	

Introduction to Benefits of Data Visualization: Data Visualization involve the graphical representation of the data in the form of charts, graph, and maps which make the analysis of the complex data easier to the user, due to this data visualization has a huge benefit in the field of Healthcare industries, Military and Finance Industry along with this it is also helpful in industrial applications like data comparison, data analysis, data distribution over a period of time and visualization of a geographical dataset.

Benefits of Data Visualization and Its Applications: Data visualization brings easiness in understanding and increases effectiveness. The human mind learns fast from visuals than that from text and tables. It is applied to a large population; for e.g., one can remember dialogues and scenes of Sholay movie, which he might have watched years before; on the other hand, it is difficult for him to recall the engineering subjects.

Nowadays, we have a good number of tools for data visualization tools, which are fast and effective. Data visualization creates a better selling strategy. Data visualization boosts the ability to process information in an easy /faster way to compare and make conclusions. For e.g., pie-charts: gives percentage-wise allocation, bar graphs: gives a better understanding of statistics.

We will discuss the Benefits of Data Visualizations only with some industries, but this applies to almost all industries.

- **Healthcare Industries**

  Creation of a dashboard to visualize patients' history can help an existing or a new doctor understand a patient's condition. In case of emergency, it could provide quicker care facilities according to disease. Data visualization can help in identifying the trend instead of going through reports of 100s of pages.

  Health care is a complex process, and most of the time is wasted in analyzing the previous reports. Data visualization gives a better selling point by increasing response time. It provides matrices by which it is easier to analyze, thus increase response time.

- **Military**

  For the military, it is a matter of life and death; it is of utmost importance to have clarity of actionable data, and to take the right action, one must have clarity of data to draw out actionable insights.

  Today, the enemy is out there in the field and is also threatening through digital warfare and cyber security. It is of utmost need to gather data from multiple sources – structured and unstructured. The amount of data is quite huge, and data visualization tools play a key role to ensure the timely delivery of proper information in the best-consolidated way possible. A better understanding of historical data provides a better prediction.

  Dynamic Data Visualization helps in understanding geography/ climate better, which helps in a better approach. Military equipment and tools' cost is quite high; with bar and pie charts, it is easy to analyze existing inventory and make the purchase as per need.

- **3. Finance Industries**

Nowadays, Data visualization tools are a must for financial sectors for exploring/explain data of related clients, understanding customer behavior, having the transparent flow of information, the efficiency of decision making, etc. Data visualization helps in creating patterns for associative firms and businesses, which helps in better investment strategy. Data visualization highlights the latest trends for better business opportunities.

Create Data Visualization for Industries

Before Visualizing data, one must know what the need for business/industry is? Data visualization helps in answering that question in an easy manner. Data analysis and visualization go hand in hand; in python, we have libraries like NumPy and pandas for data analysis and for data Visualization, we have libraries like Matplotlib, Pandas Visualization, and Seaborn. Let us understand the basic purpose of data visualization. In the end, we will see the python code for data visualization. For Comparing Data Comparison plays a very important role in quantitative analysis. The following data visualization techniques can be used.

Bar graphs (Histograms): With bar graphs, it is easy to do quantitative analysis among different categories.

Stacked charts: This has added the advantage of composite data analysis with different categories.

Composition analysis

Here we define the composition of data; the following techniques are used for visualization.

Pie charts– These are circular charts and provide percentage-wise composition.

Pyramid charts- This basically is used for the hierarchy of data.

Analysis over a period

For many instances, one needs to track the data for a period and then analyze the trends. Some visualization tools for this type of cases are as follows:

Line charts- Shows basic highs and lows of the trend.

Area charts-To show cumulative data of line chart over a period

Stock charts- Are basically used for trend analysis of stocks and market indexes over a period.

Distribution analysis of data

In most cases, we need to see the relationship between features and the effect of one feature on others. Data visualization aids with the following techniques.

Scatter Plot : Here, we represent each data point as a dot and then analyze the trend.

Box plot/ violin plot: This gives a better understanding of data for outliers with the median as a center of a distribution.

Heat map charts : These are cool plots where you can show correlation and distribution with colors.

Visualization of Geographical Dataset

In many cases, we deal with geographical data, like population growth in different regions. Effect of any natural calamity in certain areas. The following graphs can help us visualize the location-wise effect.

Choropleth maps— variation across different locations.

Dot maps— same as choropleth with dots showing data density.

DRAW INSIGHTS OUT OF VISUALIZATION TOOL

- **Presenting data visually:** Visualizations can help people understand and interpret data, and can be used to create dashboards and reports.
- **Communicating complex ideas:** Visualizations can help you communicate complex ideas and help your audience see the big picture.
- **Telling data stories:** You can combine data, visualizations, and narrative to tell a story about what the data means.
- **Comparing datasets:** Dashboards can display multiple graphs and charts side by side, making it easy to compare datasets.
- **Presenting data to non-technical audiences:** Visualizations can help employees or business owners present data to non-technical audiences. Some insights you can get from visualizations include: Average, Predictive strength, Meaningful differences, Fit line, and Most frequent. Some big data visualization tools include:
  - Microsoft Power BI
  - Tableau
  - SAS Visual Analytics
  - Qlik Sense
  - Google Data Studio
  - Splunk
  - D3
  - Apache Superset
  - Google Charts
  - Grafana

When using data visualization tools, you can consider things like:

- Defining a clear purpose
- Knowing your audience
- Keeping visualizations simple
- Choosing the right visual
- Making sure your visualizations are inclusive
- Providing context
- Making it actionable

PRODUCT IMPLEMENTATION

Effective visualization for big data techniques need to go beyond just painting pretty pictures for management. Experts say enterprises can improve their results by considering layout, designing iteratively, engaging users and understanding business needs.

- **Keep the user in mind:** Use color, form, size and placement to inform the design and use of your visualization, said Dan Gastineau, visual analytics practice lead at Aspirent, an Atlanta-based management consulting firm.

  Aspirent uses color to draw attention to the aspects of the analysis that it wants users to focus on. Size effectively communicates quantity, but too much use of different sizes to communicate information can become confusing. Instead, size is employed selectively in places where consultancy team members want to emphasize a point. Form determines the shapes used to present an analysis: for example, whether to use line or bar charts to present certain types of information. Placement of objects is as critical to effective communication as the objects themselves, Gastineau said.
- **Tell a coherent story:** Speak to your audience and keep design simple and focused. Minute details like colors to the number of charts can help ensure that a dashboard tells a coherent story. "A dashboard, much like a book, needs design elements that keep the reader in mind and does not simply force fit all the data one has access to," said Saurabh Abhyankar, senior vice president of product management at MicroStrategy. The design of dashboards will be a factor that drives adoption.
- **Prepare to design iteratively:** Work in ways to elicit frequent feedback from visual analytics users. Data exploration sparks new ideas and questions over time, and making it more pertinent over time and over adoption makes users smarter.

Solicit and incorporate feedback from your recipients to improve the experience. Building a quick proof of concept, getting feedback quickly and iterating tends to lead to a better result, faster, said Nick Mihailovski, lead product manager for Data Studio at Google Cloud. Even incorporating surveys and forms into polished reports can help ensure that the result of visualization for big data efforts indeed aids the intended recipient.

- **Personalize everything:** Make sure that the dashboard reveals personalized information to the end user, and make it relevant. Ensuring that the visualizations are responsive in design to the devices they're on and offering offline access to end users will take it a long way. Engage your audience, and propagate a data culture by using well-designed and interactive visualizations to make analytics engaging and fun, Mihailovski said. It should also be intuitive for employees to access, visualize and share their reports with live, dynamic data.
- **Start with the analysis objective:** Ensure that the data type and analysis objective informs what visualization type is chosen. "People often take a backwards approach by seeing a neat or obscure visualization type and then trying to fit their data to it," Mihailovski said. A simple table or bar chart may sometimes be most effective for visualization for big data projects.
- **Keep governance in mind:** This might take work, but it's important that end users trust the data. Gather all the help you need from a technology, process and people standpoint to ensure that the data is vetted and accurate, Abhyankar said.
- **Empathize with the viewer:** Every situation invites a different approach to visualization. For example, many data visualization gurus categorically forbid the use of pie charts , since the human eye and mind can much more readily measure differences between lengths or positions than between angles. But turning a pie chart into a bar chart can make it less obvious that the different parts add up to a single whole. It is also harder to recognize certain fractions. Consider variations like a donut chart, which is just a pie chart missing the middle for quickly showing patterns like a 75%, 20% and 5% breakdown.
- **Understand the business:** Spend some time with business users to understand what they want to achieve from the visualization for big data product, and what data needs to be used to provide them with the required insights. If necessary, invest in some tools or

techniques to profile and transform the data. "We are dealing with tons of data in big data hence, empowering the user to gain from this scale is important," said Naresh Agarwal, head of data and analytics at Brillio, a technology consultancy.

Knowing the trends in the business is important to help users apply the latest metrics and analytics to drive better business decisions. Different dashboards need to be conceived, keeping the end user in mind. Management, analysts, IT and business users will derive value from different types of visual analytics explorations.

Visualization for big data is useful only if it addresses the interests of the stakeholder for whom the analysis is performed. Clearly understanding stakeholder motivations will enable the development of visualizations that are at the right level of granularity and detail to allow stakeholders to act upon them, said Nitin Bajaj, director of business intelligence and analytics at NTT Data Services.

- **Connect visualizations**: Ensure that different visualizations on the dashboard are connected and can be linked instantly to complete the story. For example, if you are analyzing a locationwise sales summary report, you should also be able to do a year-on-year analysis or compare sales across different products, said Pratik Jain, technical architect at Kyvos Insights, a business intelligence software provider.

Make sure visualizations for big data can be updated and interrogated in real time. A static display, or one that doesn't reveal where its underlying data comes from, is not helpful in analyzing the fast-moving, varied streams of big data that enterprises must process, said Zachary Jarvinen, senior analytics product marketing manager at OpenText, a supply chain software provider.

- **Simplify where possible:** Most leading visualization for big data tools are feature-rich, which often results in a tendency by analysts to build visualizations that are dense and overly complex. This can make it difficult to glean actionable insights. A good analyst will simplify visualizations to address only the hypothesis at hand, Bajaj said. Then the analyst should check in with the stakeholder early and often to ensure that the final product is not loaded with bells and whistles that look cool on first blush, but do not address the needs of the stakeholder directly.