

Data Visualization

Data visualization is actually a set of data points and information that are represented graphically to make it easy and quick for user to understand. Data visualization is good if it has a clear meaning, purpose, and is very easy to interpret, without requiring context. Tools of data visualization provide an accessible way to see and understand trends, outliers, and patterns in data by using visual effects or elements such as a chart, graphs, and maps.

Characteristics of Effective Graphical Visual:

- It shows or visualizes data very clearly in an understandable manner.
- It encourages viewers to compare different pieces of data.
- It closely integrates statistical and verbal descriptions of data set.
- It grabs our interest, focuses our mind, and keeps our eyes on message as human brain tends to focus on visual data more than written data.
- It also helps in identifying area that needs more attention and improvement.
- Using graphical representation, a story can be told more efficiently. Also, it requires less time to understand picture than it takes to understand textual data.

Categories of Data Visualization ;

Data visualization is very critical to market research where both numerical and categorical data can be visualized that helps in an increase in impacts of insights and also helps in reducing risk of analysis paralysis. So, data visualization is categorized into following categories:

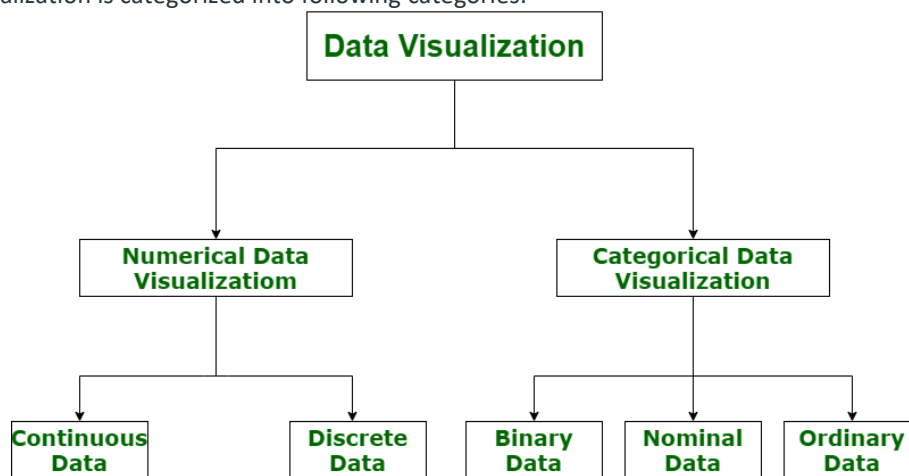


Figure – Categories of Data Visualization

1. Numerical Data :

Numerical data is also known as Quantitative data. Numerical data is any data where data generally represents amount such as height, weight, age of a person, etc. Numerical data visualization is easiest way to visualize data. It is generally used for helping others to digest large data sets and raw numbers in a way that makes it easier to interpret into action. Numerical data is categorized into two categories:

- **Continuous Data –**
It can be narrowed or categorized (Example: Height measurements).
- **Discrete Data –**

This type of data is not “continuous” (Example: Number of cars or children’s a household has).

The type of visualization techniques that are used to represent numerical data visualization is Charts and Numerical Values. Examples are Pie Charts, Bar Charts, Averages, Scorecards, etc.

2. Categorical Data :

Categorical data is also known as Qualitative data. Categorical data is any data where data generally represents groups. It simply consists of categorical variables that are used to represent characteristics such as a person’s ranking, a person’s gender, etc. Categorical data visualization is all about depicting key themes, establishing connections, and lending context. Categorical data is classified into three categories:

- **Binary Data –**
In this, classification is based on positioning (Example: Agrees or Disagrees).
- **Nominal Data –**
In this, classification is based on attributes (Example: Male or Female).
- **Ordinal Data –**

In this, classification is based on ordering of information (Example: Timeline or processes).

The type of visualization techniques that are used to represent categorical data is Graphics, Diagrams, and Flowcharts. Examples are Word clouds, Sentiment Mapping, Venn Diagram, etc.

What is Data Visualization and Why is It Important?

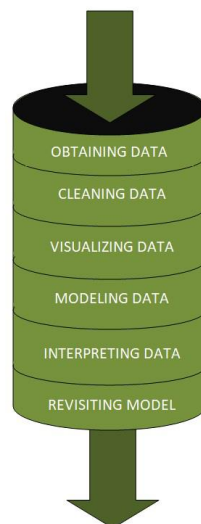
Before jumping into the term “Data Visualization”, let’s have a brief discussion on the term “Data Science” because these two terms are interrelated. But how? Let’s understand. So, in simple terms, **“Data Science is the science of analysing raw data using statistics and machine learning techniques with the purpose of drawing conclusions about that information”**. But do you know what is **Data Science Pipeline**?

In simple words, a pipeline in data science is “a set of actions which changes the raw (and confusing) data from various sources (surveys, feedback, list of purchases, votes, etc.), to an understandable format so that we can store it and use it for analysis.”



Data science pipeline in a simplified way

The raw data undergoes different stages within a pipeline, which are:



1. Fetching/Obtaining the Data
2. Scrubbing/Cleaning the Data
3. **Data Visualization**
4. Modeling the Data
5. Interpreting the Data
6. Revision

So now we are jumping to the **Data Visualization term**.

Data visualization is the graphical representation of information and data in a pictorial or graphical format (Example: charts, graphs, and maps). Data visualization tools provide an accessible way to see and understand trends, patterns in data, and outliers. Data visualization tools and technologies are essential to analysing massive amounts of information and making data-driven decisions. The concept of using pictures is to understand data that has been used for centuries. General types of data visualization are Charts, Tables, Graphs, Maps, Dashboards. Data visualization is very critical to market research where both numerical and categorical data can be visualized, which helps in an increase in the impact of insights and also helps in reducing the risk of analysis paralysis. So, data visualization is categorized into the following categories:

Let's now discuss some of the Advantages of Data Visualization.

Advantages of Data Visualization

- 1. Better Agreement:** In business, for numerous periods, it happens that we need to look at the exhibitions of two components or two situations. A conventional methodology is to experience the massive information of both the circumstances and afterward examine it. This will clearly take a great deal of time.
- 2. A Superior Method:** It can tackle the difficulty of placing the information of both perspectives into the pictorial structure. This will unquestionably give a superior comprehension of the circumstances. For instance, Google patterns assist us with understanding information identified with top ventures or inquiries in pictorial or graphical structures.
- 3. Simple Sharing of Data:** With the representation of the information, organizations present another arrangement of correspondence. Rather than sharing the cumbersome information, sharing the visual data will draw in and pass on across the data which is more absorbable.
- 4. Deals Investigation:** With the assistance of information representation, a salesman can, without much of a stretch, comprehend the business chart of items. With information perception instruments like warmth maps, he will have the option to comprehend the causes that are pushing the business numbers up just as the reasons that are debasing the business numbers. Information representation helps in understanding the patterns and furthermore, different variables like sorts of clients keen on purchasing, rehash clients, the impact of topography, and so forth.
- 5. Discovering Relations Between Occasions:** A business is influenced by a lot of elements. Finding a relationship between these elements or occasions encourages chiefs to comprehend the issues identified with their business. For instance, the online business market is anything but another thing today. Each time during certain happy seasons, like Christmas or Thanksgiving, the diagrams of online organizations go up. Along these lines, state if an online organization is doing a normal \$1 million business in a specific quarter and the business ascends straightaway, at that point they can rapidly discover the occasions compared to it.
- 6. Investigating Openings and Patterns:** With the huge loads of information present, business chiefs can discover the profundity of information in regard to the patterns and openings around them. Utilizing information representation, the specialists can discover examples of the conduct of their clients, subsequently preparing for them to investigate patterns and open doors for business.

Disadvantages of data visualization

- 1. Can be time-consuming:** Creating visualizations can be a time-consuming process, especially when dealing with large and complex datasets. This can slow down the machine learning workflow and reduce productivity.
- 2. Can be misleading:** While data visualization can help identify patterns and relationships in data, it can also be misleading if not done correctly. Visualizations can create the impression of patterns or trends that may not actually exist, leading to incorrect conclusions and poor decision-making.
- 3. Can be difficult to interpret:** Some types of visualizations, such as those that involve 3D or interactive elements, can be difficult to interpret and understand. This can lead to confusion and misinterpretation of the data.
- 4. May not be suitable for all types of data:** Certain types of data, such as text or audio data, may not lend themselves well to visualization. In these cases, alternative methods of analysis may be more appropriate.
- 5. May not be accessible to all users:** Some users may have visual impairments or other disabilities that make it difficult or impossible for them to interpret visualizations. In these cases, alternative methods of presenting data may be necessary to ensure accessibility.

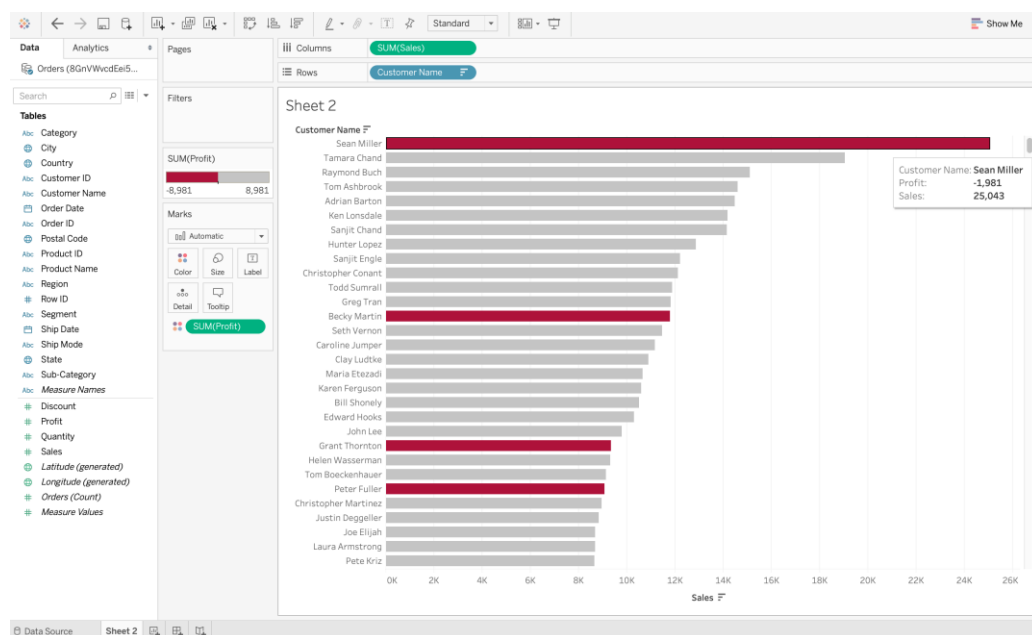
Now the most important question arises.

Why is Data Visualization So Important?

Let's take an example. Suppose you compile a data visualization of the company's profits from 2010 to 2020 and create a line chart. It would be very easy to see the line going constantly up with a drop in just 2018. So you can observe in a second that the company has had continuous profits in all the years except a loss in 2018. It would not be that easy to get this information so fast from a data table. This is just one demonstration of the usefulness of data visualization. Let's see some more reasons why data visualization is so important.

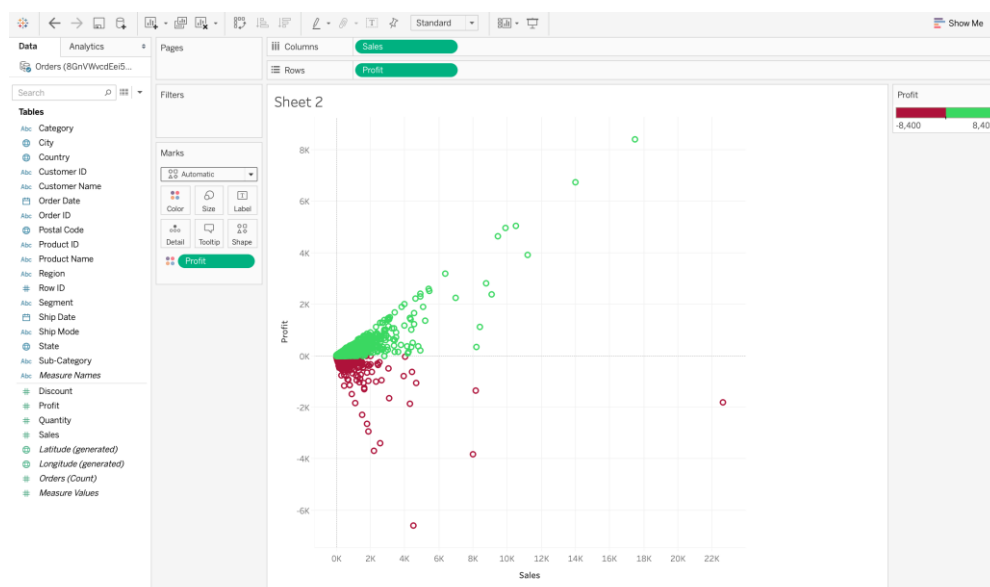
1. Data Visualization Discovers the Trends in Data

The most important thing that data visualization does is discover the trends in data. After all, it is much easier to observe data trends when all the data is laid out in front of you in a visual form as compared to data in a table. For example, the screenshot below on Tableau demonstrates the sum of sales made by each customer in descending order. However, the color red denotes loss while grey denotes profits. So it is very easy to observe from this visualization that even though some customers may have huge sales, they are still at a loss. This would be very difficult to observe from a table.



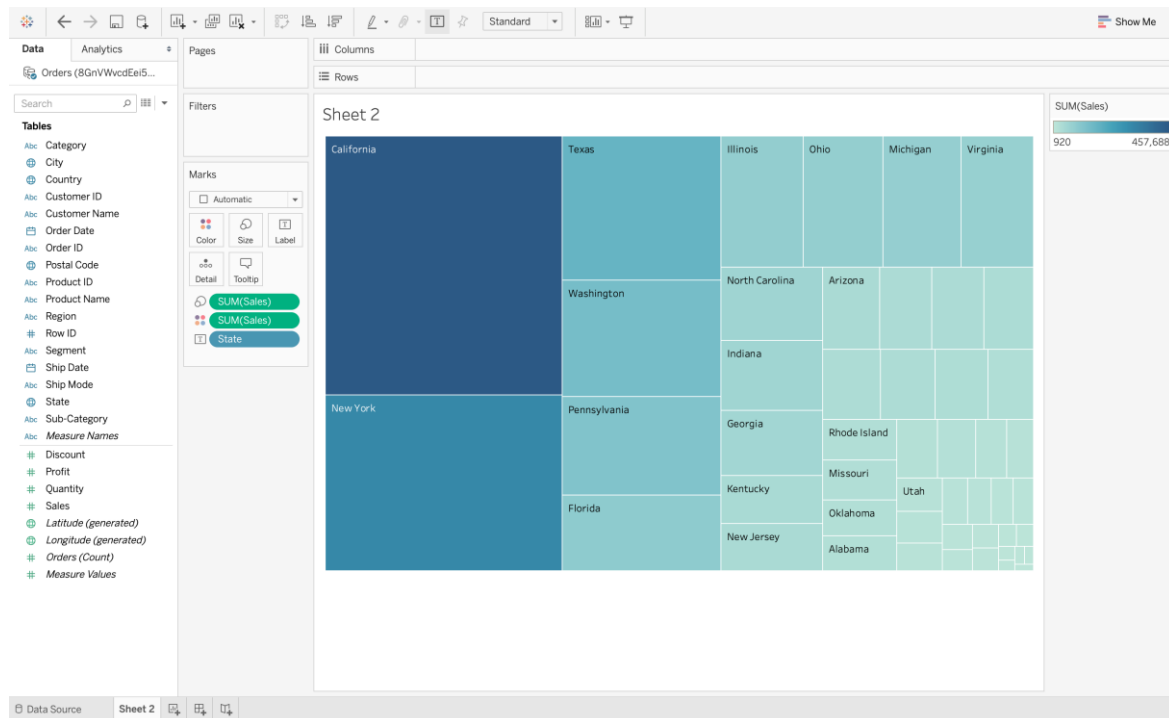
2. Data Visualization Provides a Perspective on the Data

Data Visualization provides a perspective on data by showing its meaning in the larger scheme of things. It demonstrates how particular data references stand with respect to the overall data picture. In the data visualization below, the data between sales and profit provides a data perspective with respect to these two measures. It also demonstrates that there are very few sales above 12K and higher sales do not necessarily mean a higher profit.



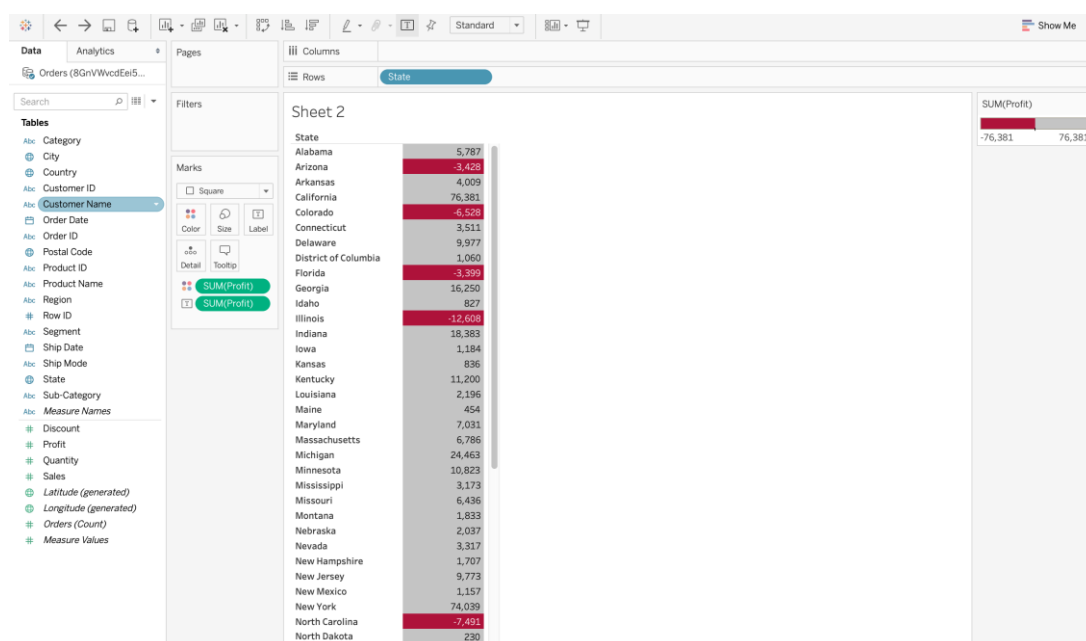
3. Data Visualization Puts the Data into the Correct Context

It is very difficult to understand the context of the data with data visualization. Since context provides the whole circumstances of the data, it is very difficult to grasp by just reading numbers in a table. In the below data visualization on Tableau, a TreeMap is used to demonstrate the number of sales in each region of the United States. It is very easy to understand from this data visualization that California has the largest number of sales out of the total number since the rectangle for California is the largest. But this information is not easy to understand outside of context without data visualization.



4. Data Visualization Saves Time

It is definitely faster to gather some insights from the data using data visualization rather than just studying a chart. In the screenshot below on Tableau, it is very easy to identify the states that have suffered a net loss rather than a profit. This is because all the cells with a loss are colored red using a heat map, so it is obvious states have suffered a loss. Compare this to a normal table where you would need to check each cell to see if it has a negative value to determine a loss. Obviously, data visualization saves a lot of time in this situation!



5. Data Visualization Tells a Data Story

Data visualization is also a medium to tell a data story to the viewers. The visualization can be used to present the data facts in an easy-to-understand form while telling a story and leading the viewers to an inevitable conclusion. This data story, like any other type of story, should have a good beginning, a basic plot, and an ending that it is leading towards. For example, if a data analyst has to craft a data visualization for company executives detailing the profits on various products, then the data story can start with the profits and losses of various products and move on to recommendations on how to tackle the losses.

To find out more points please refer to this article: [Why is Data Visualization so Important?](#)

Top Data Visualization Tools

The following are the 10 best Data Visualization Tools

1. Tableau
2. Looker
3. Zoho Analytics
4. Sisense
5. IBM Cognos Analytics
6. Qlik Sense
7. Domo
8. Microsoft Power BI
9. Klipfolio
10. SAP Analytics Cloud

To find out more about these tools please refer to this article: [Best Data Visualization Tools](#)

Top Data Visualization Libraries Available in Python, R, and JavaScript

The following are the top Data Visualization Libraries

- **Python:**
 - [Matplotlib](#)
 - [Plotly](#)
 - ggplot
 - [Seaborn](#)
 - Altair
 - Geoplotlib
 - [Bokeh](#)
- **R:**
 - ggplot2
 - Plotly
 - Leaflet
 - Esquisse
 - Lattice
- **Javascript:**
 - D3.js
 - Chart.js
 - Plotly

To find out more about these libraries please refer to this article: [Top Libraries for Data Visualization](#)