



Lecture 2: Data types and charts

Dataset:

https://docs.google.com/spreadsheets/d/13eUPgmPKSmpCs_mHNVRZWTGp98zhBFt_HBd2qQQZy1Q/edit#gid=120953924

Introduction to Data Types:

Data types in Power BI refer to the categorization of values within a dataset, specifying the kind of information each column contains. Properly defining data types is crucial for accurate analysis and effective visualization. Power BI provides various data types, and choosing the right one ensures that the software interprets and presents the data correctly.

Common Data Types in Power BI:

- **Text (String):**
 - Represents alphanumeric characters.
 - Used for fields containing names, descriptions, or any textual data.
- **Whole Number (Integer):**
 - Represents whole numbers without decimal places.
 - Ideal for counting or tracking discrete items.
- **Decimal Number (Decimal):**
 - Represents numbers with decimal places.
 - Suitable for any numeric data requiring precision, such as financial figures.
- **Date/Time:**
 - Represents dates and times.
 - Facilitates time-based analysis and visualisation.
- **Boolean/Binary:**
 - Represents true/false values.
 - Useful for binary data such as yes/no, true/false or conditions.

Charts In PowerBI:

Bar Charts

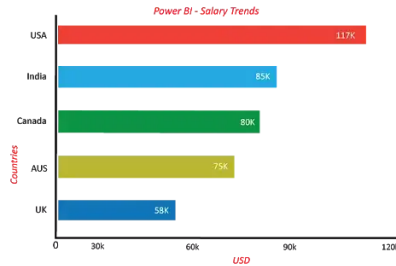
In the list of Power BI visualisation types, next, we are going to discuss bar charts.

Bar charts are mostly used graphs because they are simple to create and easy to understand. Bar charts are also called horizontal charts that represent the absolute data. They are useful to display the data that include negative values because it is possible to position the bars above and below the x-axis.

For Instance:

We have shown you the [Power BI Developer Salary trends](#) (Based on neuvvo.com) in different countries using the bar chart.

The above image shows the comparison of Power BI developer Salary trends in 5 different countries (UK, India, Canada, Australia, USA).



Why they are used :-

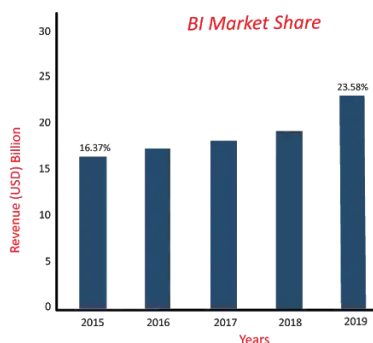
1. Bar charts are straightforward and easy to understand. They present data in a simple format, making them accessible to a wide audience, including those with non-technical backgrounds
2. Bar charts are suitable for representing absolute values, making them ideal for showcasing specific quantities, such as Power BI developer salaries in different countries.
3. Bar charts are versatile in handling data that includes negative values. By positioning bars above and below the x-axis, they can effectively display both positive and negative values, providing a comprehensive view of the data

Column Charts

Column charts are similar to bar charts, and the only difference between these two is, column charts divide the same category data into the clusters and compare within the clusters. Also, it compares the data from other clusters.

For Instance:

Let us consider one example in which we compared the BI market share with the past years.

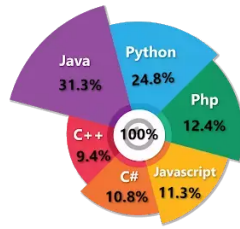


Why they are used :-

1. Column charts are effective for displaying absolute values, just like bar charts. Each column's height directly corresponds to the magnitude of the data it represents, making it easy to compare values across different categories.
2. Column charts are particularly useful when comparing values across different categories or groups. Each column typically represents a distinct category, and the relative heights of the columns facilitate quick comparisons.
3. Power BI allows for the creation of stacked column charts, where multiple data series are stacked on top of each other within a single column. This can be useful for showing the total and its components.

Pie Charts

A pie chart is a circular statistical chart, and it shows the whole data in parts. Each portion of a pie chart represents the percentages, and the sum of all parts should be equal to 100%. The whole data can be divided into slices to show the numerical propositions of each part of the data. Pie charts are mostly used to represent the same category of data. It helps users to understand the data quickly. They are widely used in education, the business world, and communication media.

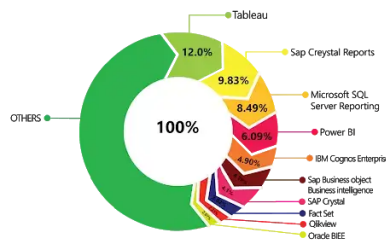


Why they are used :-

1. Pie charts are effective for illustrating the proportions or percentages of different parts in relation to a whole. Each "slice" of the pie represents a distinct category or component, and the entire pie represents the total
2. Pie charts make it easy to see the relative contribution of each category to the total. The size of each slice is proportional to the percentage it represents in the overall composition

Doughnut Charts

Doughnuts are similar to pie charts, and it is named doughnut chart because it looks similar to a doughnut. You can easily understand the data because doughnut charts show the whole data into the proposition. It is the most useful chart when you need to display various propositions that make up the final value.

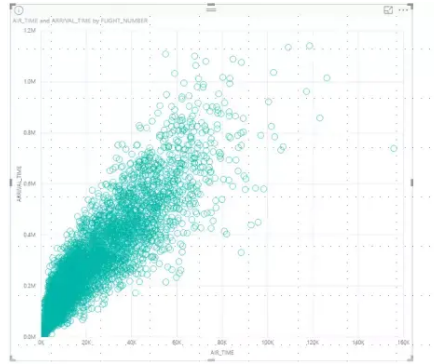


Why they are used :-

1. Similar to pie charts, doughnut charts are effective for illustrating the parts of a whole. They display the relative proportions or percentage of different categories in relation to the entire dataset.
2. Doughnut charts, like pie charts, are relatively simple and easy to understand. The ring structure maintains the intuitive representation of proportions, making it suitable for conveying high-level insights to a broad audience
3. Users can add data labels and legends to doughnut charts in Power BI, providing information about the values associated with each segment

Scatter Charts

Scatter charts are used to visualize the data using the dots that represent the values obtained from two different variables, such as the x-axis and y-axis. These charts are used to show the relationship between two different variables. It is also called a correlation plot because it shows how two variables are correlated to each other.



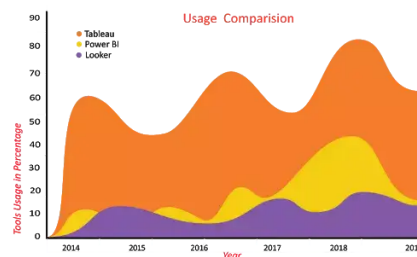
Why they are used :-

1. Scatter charts allow for a quick comparison of individual data points. It's easy to see how each point relates to others, providing a comprehensive view of the distribution and dispersion of the data.
2. Scatter charts are particularly useful for exploring and visualizing relationships between two continuous variables. Each point on the chart represents a data point with values for both variables, allowing you to identify patterns, trends, or correlations.
3. Scatter charts help in identifying clusters or patterns within the data. If certain groups or trends exist, they can be visually discerned based on the distribution of points on the chart.
4. Outliers, or data points that significantly differ from the general pattern, can be easily identified in a scatter chart. This is crucial for understanding the variability and potential anomalies in the dataset.

Area Charts

The area chart depends on line charts to display quantitative graphical data. The area between the axis and lines is commonly filled with colors, textures, and patterns. You can compare more than two quantities with area charts. It shows the trend changes over time and can be used to attract the attention of the users to know the total changes across the trends.

For Instance: The below Area chart clearly shows you how the usage of Tableau, Power BI, and Looker varies over the past six years.



Why they are used:-

1. Area charts are often used to display the distribution of values over time. The filled area between the line and the x-axis provides a visual representation of the cumulative or stacked values, making it easy to observe trends and patterns.
2. Area charts excel at illustrating cumulative data, where each data point represents the sum of the previous values. This is particularly useful for showcasing trends and the overall growth or decline in a dataset.
3. Area charts can be used to compare multiple categories simultaneously. When different areas are stacked or overlapped, it becomes easy to see the relative contributions of each category to the whole.

Line Charts

Line charts are mostly used charts to represent the data and are characterized by a series of data points connected by a straight line. Each point in the line corresponds to a data value in the given category. It shows the exact value of the plotted data. Line charts should only be used to measure the trends over a period of time, e.g. dates, months, and years

For Instance:

The below line chart shows the popularity of the Microsoft Power BI keyword in Google search across the world. It's clearly indicating that the popularity of Power BI has been increasing gradually since its inception.



Why they are used:-

1. Line charts are particularly well-suited for displaying trends and patterns in data. The continuous line connecting data points helps visualize the direction and magnitude of changes over time or other continuous variables.
2. Line charts are often used to visualize time-series data, where the x-axis represents time intervals. This makes them ideal for tracking changes and identifying patterns over different time periods
3. The connected line in a line chart visually emphasizes the continuity and connection between data points. This is especially useful when the order of data points is meaningful, such as in time-based or sequential data.

Combo Charts

A combo chart is a combination of both the column charts and line charts that help you to make a quicker comparison of the data. The combo chart shows the relationship between two measures in a single visualization. It also helps to compare multiple measures with different values.



Why they are used:-

1. Combo charts are useful when you want to visualize and compare different types of data that have distinct scales. For example, you might want to show both sales revenue (numeric values) and the number of units sold (count data) on the same chart.
2. The most common type of combo chart combines a line chart with a column or bar chart. The line chart is typically used to represent trends or continuous data, while the column or bar chart is used for categorical or discrete data
3. Combo charts often feature dual axes, where each chart type has its own y-axis. This allows for better comparison of data with different units or scales, as each data series can be plotted against its corresponding axis.

Maps

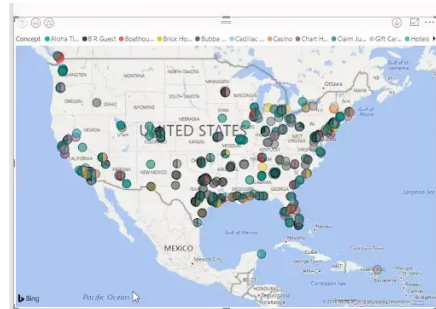
Maps are divided into three types, and they are listed below:

- Regional Maps
- Point Maps

- Flow Maps

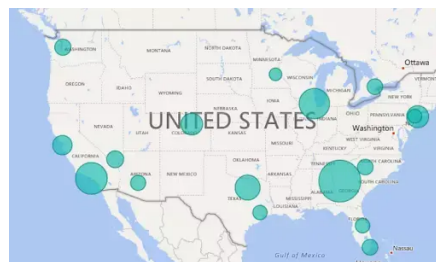
A) Regional Maps:

Regional maps use different colors to represent the distribution of a specific range of values on the map.



B) Point Maps

A point map is used to represent the geographical distribution of data by plotting the same size points on the geographical background. It helps the user to grasp the overall distribution of the data, but it is a tough task if you want to observe specific data.



C) Flow Maps

A flow map is a type of map that is particularly designed to show specific themes connected with a particular geographical area. It is used in cartography to show the movements of the objects between two or more areas.



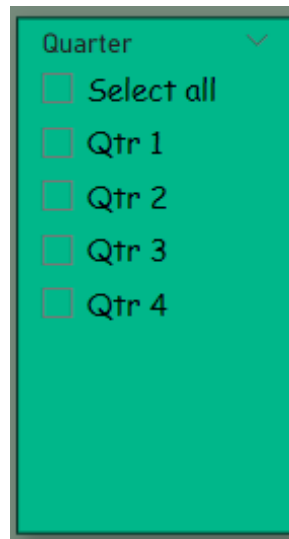
Why they are used:-

1. Maps enable location-based analysis, helping users understand how data is distributed across different areas. This is crucial for tasks such as regional sales analysis, customer distribution, and supply chain management.

2. Marker maps display data points as markers on a map, each associated with a specific location. This is useful for representing individual data points, such as customer locations, store branches, or other geospatially relevant entities

Slicers Charts

Slicers charts are visual filters. Using slicers, you can filter or sort your data by clicking on the type of data you want. In the below example, you can see all-region sales. In case if you want to see a particular region's sales, then click on that region, and it shows the specific region's sales.

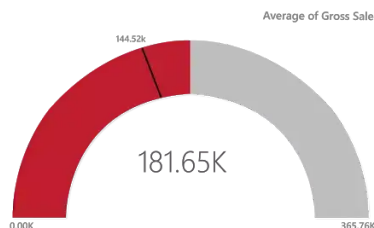


Why they are used:-

1. Slicers enable users to interactively filter data in charts. When users select a value or range in a slicer, the associated charts on the report update to reflect the chosen criteria.
2. Slicers are typically connected to specific fields in the dataset. For example, you might have a slicer linked to a "Region" field, allowing users to filter the data based on different regions
3. Slicers can be presented in various formats, including dropdown lists, horizontal lists, or even as a range slider. The format chosen depends on the type of data and the user experience preferences

Gauge Charts

A gauge chart is also known as a speedometer or dial chart. It uses the needle to read the data, and it shows the information on a dial. The gauge chart represents the value of each needle as it reads the data according to the axis or colored data. These charts are useful to compare the values between the variables either by using multiple needles on the same gauge or different gauges.



Why they are used:-

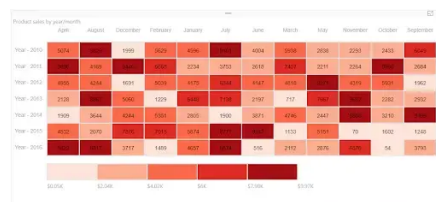
- Waterfall charts are useful for conducting bridge analyses, where each block represents a factor contributing to the change between two values. This is often employed to explain variances in financial or performance metrics
- Waterfall charts are valuable for visualising the components of profit and loss statements. The chart can show how various revenue and expense categories contribute to the net profit or loss.

Tree Maps

Treemaps display hierarchical data set in a nested rectangle. At each level, hierarchy is represented by a colour. The size of the space in the rectangle depends on the data values. The rectangular boxes are arranged in size from top left to bottom right.

For Instance:

For example, you are analysing your sales, and you have top-level branches for clothes categories: Rural, Mix, Youth, and Urban. Power BI treemaps split your categories into rectangle boxes (leaves); these boxes would be shaped and sized based on the number of sold.



Why they are used:-

- Tree maps are effective for visualising hierarchical data structures, where each rectangle represents a category or node, and the size of the rectangles corresponds to a quantitative value
- The nested rectangles in a tree map represent parent-child relationships in the hierarchy. Larger rectangles contain smaller ones, allowing users to understand the hierarchical structure and the distribution of values within each level.
- The size of each rectangle in a tree map reflects a quantitative measure, such as sales revenue, population, or market share. Users can quickly compare the sizes of rectangles to identify the relative contributions of different categories.

Matrix Chart

A Matrix chart shows the relation between two or more variables in a data set. It is mainly made up of columns and rows to represent the data in the grid format. At least two variables are required to create a matrix chart if there is any third or fourth variable, and colour or other dimensions that can be added to the matrix to represent the data.

For Instance:

The below matrix chart represents the company's revenue in different years with other factors.

| Calendar Year | 1 | 2 | 3 | 4 | Total |
|---------------|-----------------|-----------------|-----------------|-----------------|------------------|
| 2005 | | | \$4,647,156.86 | \$6,684,652.11 | \$11,331,808.97 |
| 2006 | \$5,860,884.49 | \$6,167,832.57 | \$10,277,073.05 | \$8,368,581.06 | \$30,674,771.18 |
| 2007 | \$6,679,873.81 | \$8,357,874.87 | \$13,670,536.66 | \$13,285,444.49 | \$41,993,729.83 |
| 2008 | \$11,386,315.07 | \$14,371,806.64 | \$50,840.63 | | \$25,808,962.35 |
| Total | \$23,927,073.37 | \$28,897,514.09 | \$28,645,607.20 | \$28,339,079.67 | \$109,809,274.32 |

Why they are used:-

- Matrix charts present data in a tabular format with rows and columns, allowing users to view the intersections of various categories or dimensions
- Matrix charts support hierarchies in both rows and columns. This enables users to drill down into more detailed information by expanding or collapsing levels of the hierarchy
- Matrix charts provide the ability to aggregate and summarise data, making them useful for analysing total values, averages, counts, or other aggregated measures

- Matrix charts are often used for cross-tabulation, where the intersection of rows and columns displays the value of a particular measure. This is valuable for comparing data across different dimensions

Tables

A table is a grid that contains the related data in a series of rows and columns. Tables are useful if you are comparing the same category for many values.

For Instance:

In the below example, you can see the same category having multiple measures to compare.

| Category | This Year Sales Status | Average Unit Price | Last Year Sales | This Year Sales | This Year Sales Goal |
|-----------------|------------------------|--------------------|---------------------|---------------------|----------------------|
| 010-Womens | ● | \$7.30 | \$2,680,662 | \$1,787,958 | \$2,680,662 |
| 020-Mens | ● | \$7.12 | \$4,453,133 | \$4,452,421 | \$4,453,133 |
| 030-Kids | ● | \$5.30 | \$2,726,892 | \$2,705,490 | \$2,726,892 |
| 040-Juniors | ● | \$7.00 | \$3,105,550 | \$2,830,385 | \$3,105,550 |
| 050-Shoes | ● | \$13.84 | \$3,640,471 | \$3,574,900 | \$3,640,471 |
| 060-Intimate | ● | \$4.28 | \$955,370 | \$852,329 | \$955,370 |
| 070-Hosiery | ● | \$3.69 | \$573,604 | \$486,196 | \$573,604 |
| 080-Accessories | ● | \$4.84 | \$1,273,096 | \$1,379,259 | \$1,273,096 |
| 090-Home | ● | \$3.93 | \$2,913,647 | \$3,053,326 | \$2,913,647 |
| 100-Groceries | ● | \$1.47 | \$810,176 | \$829,776 | \$810,176 |
| Total | ● | \$5.49 | \$23,132,601 | \$22,051,952 | \$23,132,601 |

Why they are used:-

- Tables present data in a tabular format with rows and columns, providing a structured and easy-to-read display
- Tables are commonly used to display raw or detailed data. Each row represents a record, and each column represents a variable or attribute
- Tables include column headers that describe the content of each column. This helps users understand the meaning of the data and the type of information presented