**Variable**

**Numerical variable**

or

**Quantitative variable**

**Categorical variable**

or

**Qualitative variable**

**Nominal variable**

**Eg**.

**Gender** (M, F),

**Blood type** (O+, A)

**Discrete variable**

**Eg**.

**Number of books** (20,12),

**Age** (28, 55)

**Ordinal variable**

**Eg**.

**Economic status** (low income, middle income, high),

**Educational level** (BSc, MSc, PhD)

**Continuous variable**

**Eg**.

**Income** (4265.2, 5000, 826.02), **Height** (1.73,1.8), **Temperature** (37.2, 20.3)

|  |  |
| --- | --- |
| **Pie Chart**  **1 Categorical Variable + 1 numerical variable** | * Used when you want to compare the contribution of each category in % to the whole. * It’s best to arrange slices from largest to smallest. However, if the variables have a specific ordering, you should follow that. * When you have large data sets with several categories, you shouldn’t use a pie chart, as it looks crowded.      * **Excel** * Requires categorical column values to be grouped |
| **Donut chart**  **1 Categorical variable + 1 numerical variable** | * Same use as pie chart except you can display total value in the centre * Personally, I’d use pie chart if the values are % as we know the total is 100% and wouldn’t make sense to put that in the centre. Therefore, I’d use donut chart if the values are not % but rather values so that the reader can be informed about the total value in the centre of the chart.      * **Excel** * Requires categorical column values to be grouped |
| Sun burst |  |
| **Tree map**  **2 or more Categorical variables + 1 Numeric variable** | * Used when you want to visualize multiple categorical variables where there is a hierarchical relationship between them. Eg Country and City columns where Country has a higher order followed by the cities. * The tree map consists of larger rectangles having distinct colours and smaller rectangles inside each larger one. * The main larger rectangles (branches) represent the 1st order or highest-level category. The smaller rectangles inside (leaves) represent the next level order of columns and so on. The size of the rectangles is determined by the aggregate of the numeric values for the tree and branch eg. the populations, sales earned, etc. In the chart example, Ghana is the branch with Kumasi and Accra as the leaves. The sizes of Kumasi and Accra are determined by the population value. * The rectangles are arranged in size from top left (largest) to bottom right (smallest) * Highly recommended when a bar chart can't effectively handle the large number of values and when there are large amounts of hierarchical data to be displayed * **Excel** * Requires categorical column values to be grouped or scattered * Data columns should be positioned in order of hierarchy with highest level as the 1st column and so on. * The column with the highest hierarchy (1st column) should be sorted |
| **Bar chart**  **1 Categorical variable + 1 numerical variable** | * Used to compare different categorical or discrete variables, such as age groups, classes, schools * Also useful to compare different variables or show changes in data over time. * Used for comparing category values to each other * Horizontal bar chart is used to enhance label readability * Consider sorting bars by length when the alphabetical order is not necessary * **Excel** * Requires categorical column values to be grouped |
| **Stacked bar chart and 100% stacked bar chart**  **2 or more Categorical variables + 1 Numeric variable** | * Similar to bar chart except this takes more categorical variables. * The 100% stacked bar chart is used when you want to show the proportion of each category in % to the whole. * **Excel** * Requires categorical column values to be grouped |
| **Waterfall**  **1 categorical variable + 1 numerical variable** | * **NB:** Used when you want to show **changes** in the value of an entity over a period of time (eg. Transactions over a period of time). See chart example * Use colours to differentiate increases, decreases, and totals.      * **Excel** * Requires categorical column values to be grouped * **Excel Limitations:** * A change in values can appear in two ways. The new value after change can be +ve or -ve. * If expenses in Jan (5,000) reduced to 3,000 in Feb, there is a decrease, but the new value is still +ve. * Also, if expenses in Jan (5,000) reduced to -200, here there is also a decrease but this time the new value is -ve * When you create a waterfall in excel, only the second example will display the bar decrease due to the -ve value. Hence for example 1 cases, you the need to create the waterfall using the stacked bar chart approach. |
| **Funnel chart**  **1 categorical variable + 1 numerical variable** | * Funnel chart is used to visualize a process that has stages, and items flow sequentially from one stage to the next * For example, a sales funnel that tracks customers through stages: Lead > Qualified Lead > Prospect > Contract > Close * Each funnel stage represents a percentage of the total. So, in most cases, a funnel chart is shaped like a funnel -- with the first stage being the largest, and each subsequent stage smaller than its predecessor |
| **Histogram**  **1 numerical variable** | * Used when you want to summarize discrete or continuous data on an interval scale * The x-axis represents the data grouped into intervals and the y-axis represents the frequency (number) of data points in that interval |
| **Line chart**  **1 categorical variable (date) + 1 or more numerical variables** | * Line chart is used to display and compare trends over time. * In order to produce a line chart, you always need a column with date/time values. This is used as the x-axis * The line on the chart is produced by the numerical variables which are recorded for each date/time stamp. * **NB:** In the chart example we see, the lines represent variables which have a common relationship ie. they are all shops. If there is another variable we want to display which has no relationship with the lines, then we replace its line with a bars in a combo chart. More often, this variable who has no relationship with the others also has a different scale of values. Example is seen in the combo chart below |
| **Combo chart**  **1 categorical variable (date) + 2 or more numerical variables** | * Combo chart is used to compare data sets with different axis scales (eg. Sales of different companies compared with total number of shoppers). * Just like the line chart, it displays the trend over time therefore requires a date categorical variable. * **Excel** * Requires categorical column values to be grouped |
| **Ribbon chart**  **2 or more categorical variables + 1 numerical variable** | * The ribbon chart is basically a stacked column chart which rearranges the stack at each x-axis to display which data category had the highest rank (largest value) at that x-axis point followed by the next. See the graph example.      * Excel * There’s no ribbon chart in excel |
| **Area chart**  **1 categorical variable (date) + 1 or more numerical variables** | * It is basically a line chart whose area beneath the lines are filled with colour to indicate the magnitude or volume of the trend. * For example, data that represents profit over time can be plotted in an area chart to emphasize the total profit. * **NB:** This however does not mean the data points are placed exactly as they would on the line chart. The line chart would represent the data points on the chart exactly as they are recorded in the data table. Area chart on the other hand deals with the area occupied by a data point therefore the data points would be increased in certain category values (in this example Tesla sales represented in orange) to reflect the area compared to its fellow category values (Apple sales represented in blue). * Observe the data points in the two graphs. With the same data plotted on both graphs we see different shapes. * For the first area represented on the area chart (the blue area at the base),the data points are always the same as its line chart because it’s the first line or category value calculated directly from the x-axis base. So the area was calculated from the base x-axis at 0. That is why the blue highlights have the same data point representation on both graphs. Apple sales (blue) was picked by the chart as the first line or area by the excel chart in both graphs. * As there are more lines or category values included in the chart, for instance the orange highlight, we notice its area datapoint has gone further higher than in the line chart. At January, the original data point for Apple (blue) as correctly represented in the line chart is 38 and Tesla (orange) is 100. In the area chart at January, Apple (blue) is still the same 38 but Tesla (orange) has gone to around 140. Tesla (orange) at January was 100 and it needs to represent how big that area is so it adds it to where the area of Apple (blue) ended (38). So if Apple was represented at 38, it adds Tesla (orange) value of 100 on its top to let it rise to 138. * If the area datapoint had not been increased on top of Apple (blue) and rather it just used the data point of 100 just like the line chart, we wouldn’t see the full magnitude of Tesla (orange) 100 datapoint value because some portion of it would have been covered by the blue part. So it needs to add it on top so that we see its full area. * Its like comparing the height of two pencils (a shorter and longer one). If we place the longer one behind the shorter one, we wouldn’t be able to judge its height correctly since its bottom has been covered by the shorter pencil. We would be deceived to believe that its entire height is only the visible part above the shorter pencil that we see. To rectify this, we place the longer pencil on top of the shorter one to see its full height |
| Stacked area chart |  |
| **Boxplot**  **1 numerical variable**  **1 categorical + 1 numerical variable**  **2 categorical + 1 numerical variable** | * Boxplot is used when you want to summarize the numerical data for a particular column. It gives you info such as the min value in your data, the max value, the quartiles, the median, skewness of the distribution and in some cases, we can identify the mean from it. * It is also very essential when you want detect outliers in your data as they would show outside the boxes * They are efficient for a univariate chart ie. summarizing one variable (1 numerical variable) * However, they can also be used effectively to compare variables that are in the same units of measurement. |
| **Scatter plot**  **2 numerical variables**  **2 numerical + 1 categorical variable** | * While a line chart is used for comparing numerical variables which have a common relationship (eg. Tesla and Apple are both companies), a scatter chart on the other hand is good for comparing two kinds of numeric data which do not have a common relationship. (eg. Sales made and number of shoppers, or sales made at different temperatures). This is often done to identify if there is a correlation between numerical variables which do not have a common relationship (eg. Do the sales decrease when temperatures are high?). It can also be used to compare numerical variables with a common relationship. (eg. How do the sales of tesla and apple correlate, does one decrease when the other increases?) * A scatter chart is useful for distinguishing outliers, correlations, and categorization in two sets of data |
| **Bubble chart**  **1 categorical variable (filter) + 3 numerical variables (x-axis, y-axis, bubble size)** | * Used to compare numeric values * A bubble chart replaces the data points of scatter plots with bubbles, their sizes representing an additional dimension of the data. * If you want to use quadrants, or your data has three data series that each contain a set of values, choose a bubble chart. |
| Map | * Use a basic map to associate both categorical and quantitative information with spatial locations. |
| Filled map | * A filled map uses shading or tinting or patterns to display how a value differs in proportion across a geography or region. Quickly display these relative differences with shading that ranges from light (less-frequent/lower) to dark (more-frequent/more) |
| Gauge | * Gauges and single-number cards use a numeric value to indicate progress in relation to a goal. These charts work well as part of a dashboard for highlighting a single value or KPI * They are good for tracking a single data value’s progress toward a goal, such as, achievement level, revenue, savings, or number of Facebook followers or percentage over time. * A gauge chart displays a single value that measures progress toward a goal. The goal, or target value, is represented by the line (needle). Progress toward that goal is represented by the shading. And the value that represents that progress is shown in bold inside the arc. All possible values are spread evenly along the arc, from the minimum (left-most value) to the maximum (right-most value) |
| Card | * Single number cards display a single fact, a single data point such as total sales, market share year over year, or total opportunities |
| Multi-row card | * Same function as card but multi row cards display one or more data points, one per row. |
| Kpi | * Similar to gauge, it is used to communicate the amount of progress made toward a measurable goal * Provides more details by an ability to include projected metric. You can add in a trend axis which will show different colours depending on whether the metric is up or down |
| Slicer | * A slicer enables people to examine part of a data visualization more deeply through filtering. * Slicers can help people explore data like: Categories, Years, Geographical locations |
| Table |  |
| Matrix |  |
| R script visual |  |
| Python visual |  |
| Key influencers | * A key influencer chart displays the major contributors to a selected result or value * To see which factors affect the metric being analysed * Eg. what influences customers to place a second order, why were sales so high last June, do short-term contracts affect churn more than long-term contracts |
| Decomposition tree | * The decomposition tree visual lets you visualize data across multiple dimensions * It automatically aggregates data and enables drilling down into your dimensions in any order |
| Q&A | * the Q&A visual lets you ask questions about your data using natural language. |
| Smart narrative |  |
| Metrics (Preview) |  |
| Paginated report |  |
| ArcGIS Maps for Power BI |  |
| Power Apps for Power BI |  |
| Power Automate for Power BI |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |