Takaisin välilehdelle

Tehty: Käy oppitunti läpi loppuun asti

Visualizing categorical data

Where the **Table** and **Matrix** visuals allow for a detailed look at multiple measures, the visuals in this section are best for displaying a data value across multiple categories. In the upcoming visuals, you will be displaying bars, columns, and other visual elements, which will be proportional to the data value. These visuals have a far less detailed view of the data, but it is very easy and quick to distinguish the differences in the values within the chosen categories. All of the visuals allow for cross-highlighting, cross-filtering, and the use of drilldowns, which will not be a focus since it was covered in the previous examples. This section will focus on how to understand and configure the following visuals:

- Bar and column charts
- Pie and donut charts
- Treemaps
- Scatter charts

Continue using the same Power BI report from the previous examples. Start by creating a new report page called Categorical Data.

Bar and column charts

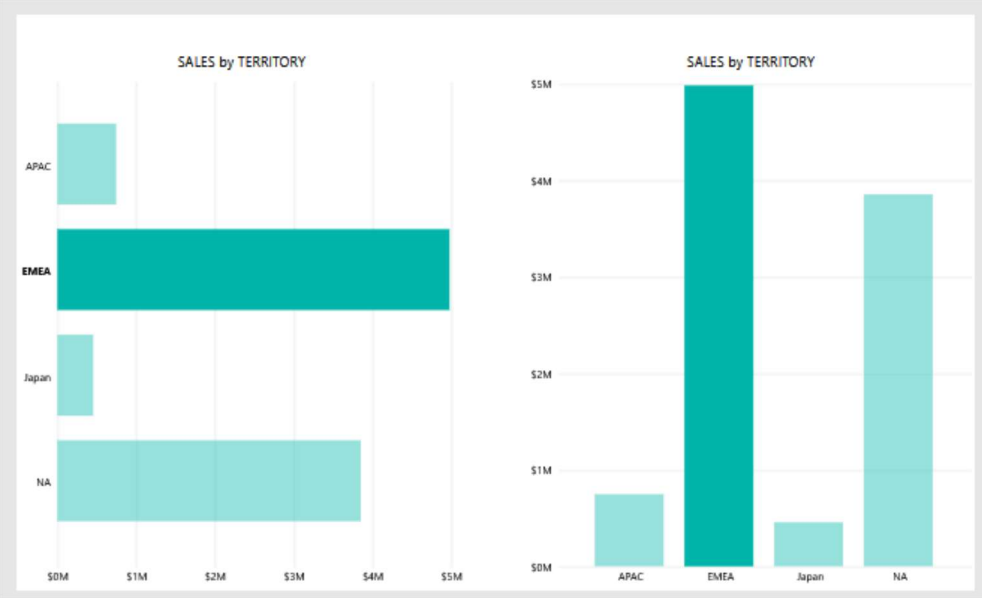
Click

to watch a video about bar and column charts.

Bar and **Column** Charts are one of the most common ways to visualize data. Both of these data visuals use rectangular bars where the size of the bar is proportional to the data values.

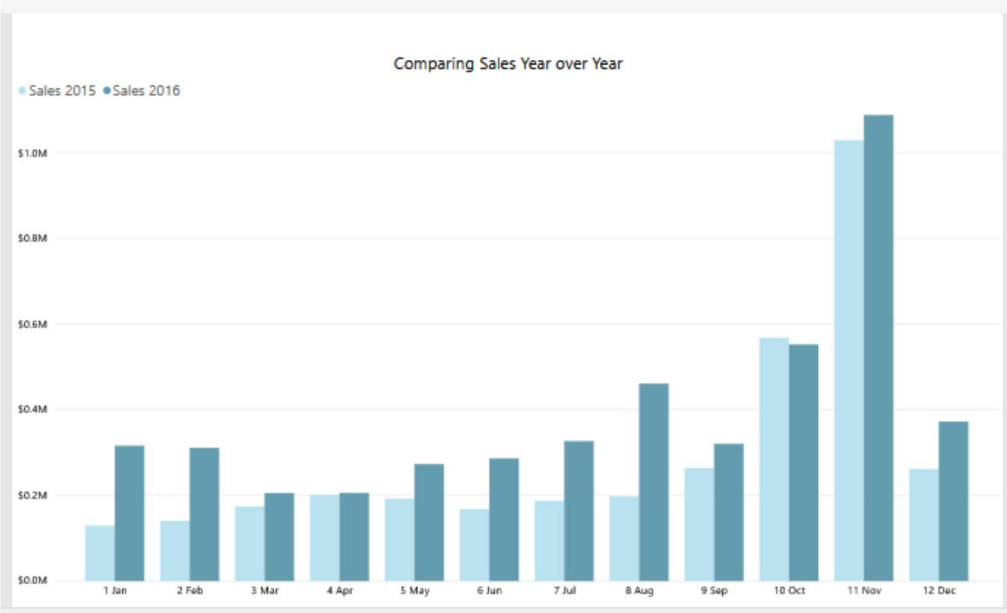
Both the **Bar** and **Column** charts are very similar in setup and how they visualize data. The only difference here will be the orientation: the **Bar chart** uses rectangular bars horizontally where the length of the bar is proportional to the amount of data, while the **Column chart** displays the bars vertically, but both are used to compare two or more values. Both visualizations have three different formats: **Stacked**, **Clustered**, and **100% stacked**.

There are some situations where the **Bar chart** will better display data, and the same thing can be said of the **Column chart**. The biggest limitation of the **Column chart** would be the limited space on the X-axis where the category would go. So, if you have a lot of data labels or if they are very long, you may find that the **Bar chart** is the better option. An example where you might choose the **Column chart** over the **Bar chart** is if your dataset contains negative values. In a **Bar chart**, the negative values will show on the left side while in a **Column chart** they will display on the bottom. Users generally associate negative values with a downward direction.



In the figure below there is a pretty standard **column chart** showing sales data over a two year period. It is recommended to seperate the two years with a different color when comparing sales between two given years, with the previous year being a lighter shade.

A **clustered column chart** is essentially a column chart that lets you compare two values side by side. In this case we are comparing sales for each month over a period of two years. Hovering over one of the values lets you see the % difference between the selected month and the sales values. A stacked column chart is very similar except the bars for each month will be stacked on top of each other rather then side by side. It is now easy to compare sales data last year vs this year for each month. For this sample data, October was the only month sales were down compare to last year.



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. It is also an artificial intelligence (AI) visualization, so you can ask it to find the next dimension to drill down into based on certain criteria. This makes it a valuable tool for ad hoc exploration and conducting root cause analysis.

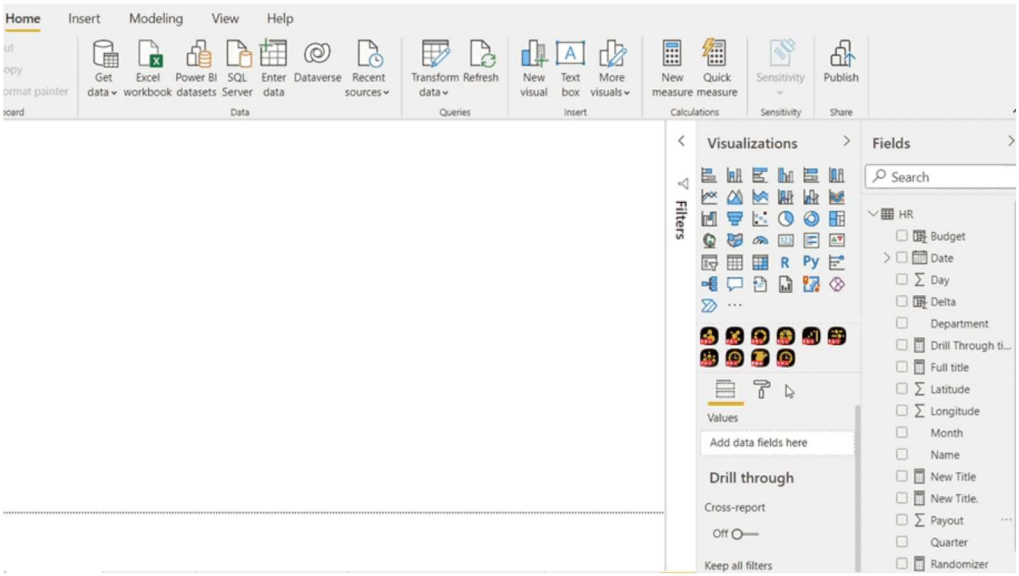
Pie and donut charts

Click

to watch a video about Power BI pie charts.

Both the **Pie chart** and **Donut chart** are meant to visualize a particular section compared to the whole, rather than comparing individual values to each other. The only difference between the two is that the **Donut chart** has a hole in the middle, which could allow for some sort of label. Both visuals can be very effective in allowing cross-highlighting, but if there are too many categories, it can become difficult to read and interpret.

Let's look at setting up a **Donut chart**:



When creating pie and donut charts, consider how filtering the data may affect the readability of the chart. In addition to having too many slices that clutter the chart, having slices that are too narrow to easily identify, or slices that are very similar in size, can be detrimental to a consumer's ability to draw accurate conclusions from the chart.

Considerations and troubleshooting

- The sum of the doughnut chart values must add up to 100%.
- Too many categories make it difficult to read and interpret.
- Doughnut charts are best used to compare a particular section to the whole, rather than comparing individual sections with each other.

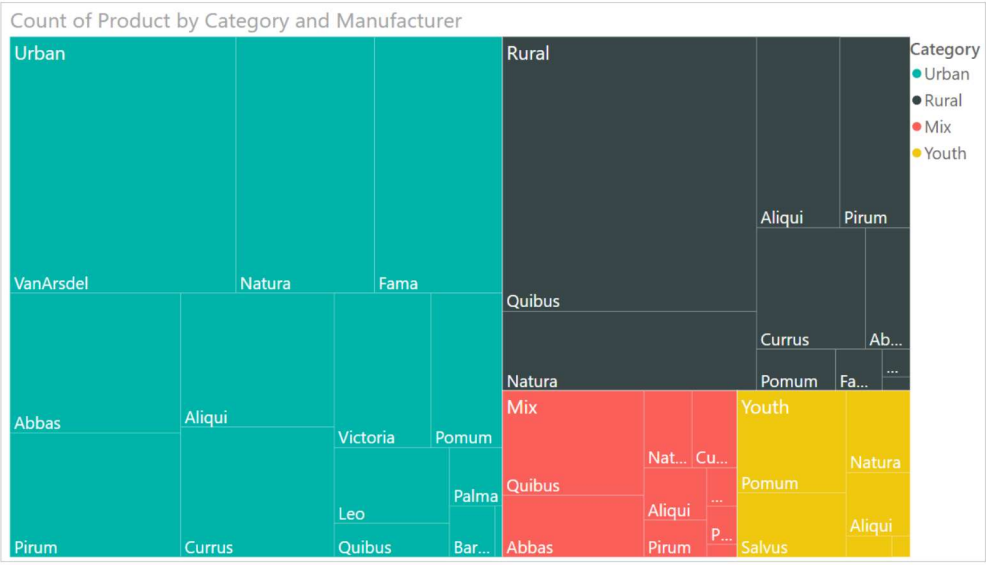
Treemap

Click

to watch video about Power BI treemaps.

Treemaps display hierarchical data as a set of nested rectangles. Each level of the hierarchy is represented by a colored rectangle (branch) containing smaller rectangles (leaves). Power BI bases the size of the space inside each rectangle on the measured value. The rectangles are arranged in size from top left (largest) to bottom right (smallest).

Look to treemaps as an alternative when bar and column charts become too cluttered by the number of categories because it still allows the consumer to compare the size of one category to another.



For example, if you're analyzing your sales, you might have top-level branches for the clothing categories: **Urban**, **Rural**, **Youth**, and **Mix**. Power BI would split your category rectangles into leaves, for the clothing manufacturers within that category. These leaves would be sized and shaded based on the number sold.

In the **Urban** branch above, lots of **VanArsdel** clothing was sold. Less **Natura** and **Fama** was sold. Only a few **Leo** were sold. So, the **Urban** branch of your Treemap has:

- The largest rectangle for **VanArsdel** in the top-left corner.
- Slightly smaller rectangles for **Natura** and **Fama**.
- Lots of other rectangles for all the other clothing sold.
- A tiny rectangle for **Leo**.

You could compare the number of items sold across the other clothing categories by comparing the size and shading of each leaf node; larger and darker rectangles mean higher value.

Treemaps are a great choice:

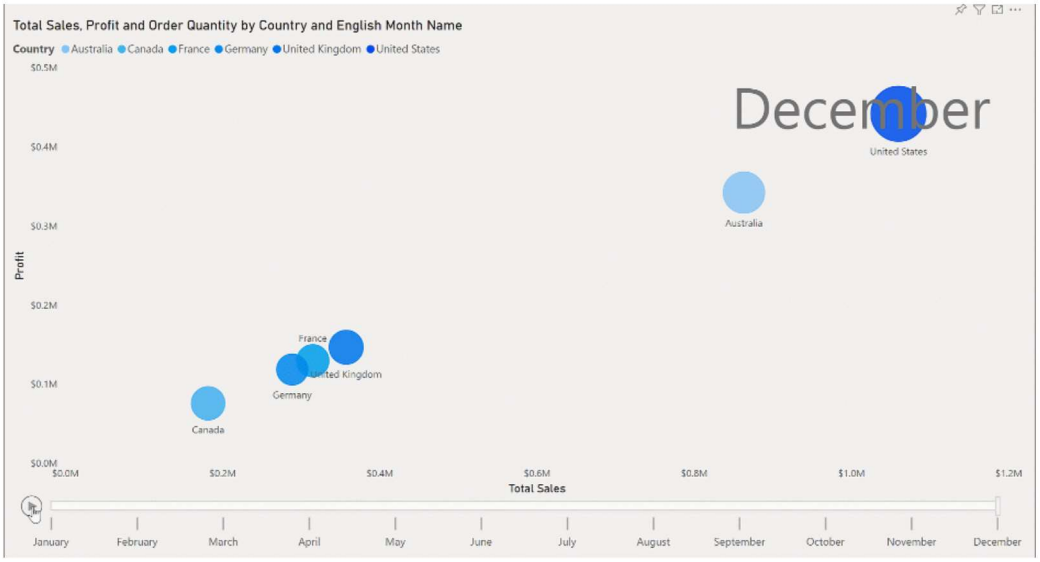
- To display large amounts of hierarchical data.
- When a bar chart can't effectively handle the large number of values.
- To show the proportions between each part and the whole.
- To show the pattern of the distribution of the measure across each level of categories in the hierarchy.
- To show attributes using size and color coding.
- To spot patterns, outliers, most-important contributors, and exceptions.

Scatter chart

Click link to watch a video about scatter charts.

The last visual used for categorical data is the **Scatter chart**, sometimes referred to as the **Bubble chart**. This visual allows you to show the relationships between two or three numerical values. You are given the opportunity to place values on the X- and Y-axes, but what is different about this visual is the ability to add a **third value for the size**, and this is where the name **Bubble chart** comes from.

There is also a very unique option available within the **Fields** section to really bring this data to life, and it is called the **Play Axis**.



Scatter charts are a great choice:

- To show relationships between two numerical values.
- To plot two groups of numbers as one series of x and y coordinates.
- To use instead of a line chart when you want to change the scale of the horizontal axis.
- To turn the horizontal axis into a logarithmic scale.
- To display worksheet data that includes pairs or grouped sets of values.
- To show patterns in large sets of data, for example by showing linear or non-linear trends, clusters, and outliers.
- To compare large numbers of data points without regard to time. The more data that you include in a scatter chart, the better the comparisons that you can make.

In addition to what scatter charts can do, bubble charts are a great choice:

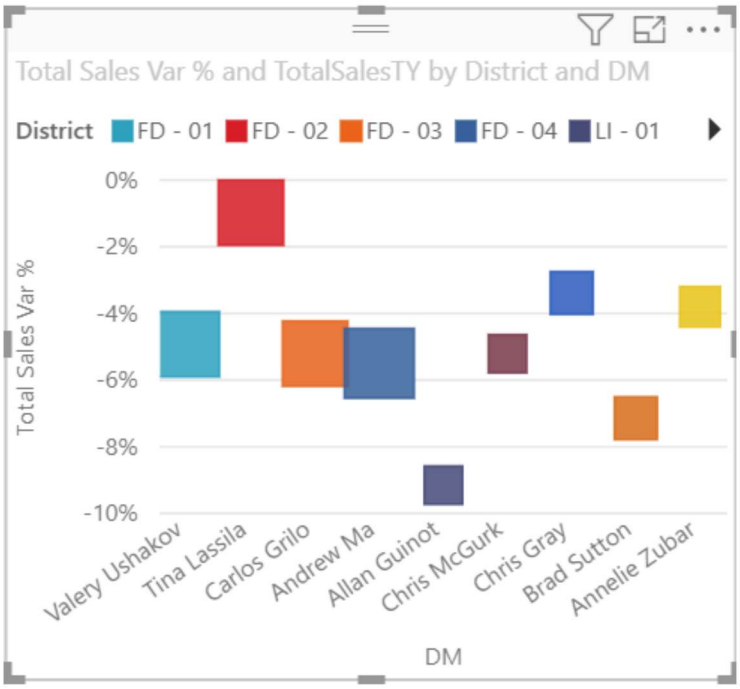
- If your data has three data series that each contains a set of values.
- To present financial data. Different bubble sizes are useful to visually emphasize specific values.
- To use with quadrants.

The **Scatter chart** uses a sampling algorithm when plotting larger sets of data to improve performance. The standard method used plots every 10^{th} data point. In general, that sampling method is adequate. However, as datasets grow larger, high-density clusters of data can be oversampled, leaving some sparse data points to be completely omitted from the chart. To solve this issue, a setting is available in the **Format** section under the **General** options, then in the **Properties** option's **Advanced options** subheader, to enable a **high-density sampling** algorithm that takes proximity to nearby data points into account, thereby ensuring the data points outside the high-density cluster are represented on the chart rather than being missed simply because they were not the 1 in 10

selected for display. This setting prioritizes an accurate distribution over accurate density on the **Scatter chart**. When showing items with no data, adding a ratio line, or using the **Play Axis**, the high-density setting will be ignored and the scatter chart will revert back to the standard sampling described at the beginning of this section.

Dot plot charts

A dot plot chart is similar to a bubble chart and scatter chart, but is instead used to plot categorical data along the X-Axis



Click

to watch a video about Power BI scatter charts.

See also Exercise 18.

Visualizing trend data

Olet suorittanut 100 % oppitunnista

100%