**Storytelling with Charts**

**Part 3: Do you want to compare items?**

This is the third article in a series aimed at helping people involved in data visualization activities to select the most appropriate chart for the type of message they are trying to show to their audience.

Three Python-based charts that allow displaying the **distribution of a single quantitative variable** were indicated in the [first article](https://medium.com/towards-data-science/storytelling-with-charts-23dd41096721) of the series.

The most appropriate charts when the message consists of **showing the magnitude of a certain set of numbers** were indicated in the [second article](https://medium.com/towards-data-science/storytelling-with-charts-dae59034f60) of the series.

Now the message is to **compare items.** The following are the charts **commonly used** for delivering such graphical representation:

* · Standard Bar Charts
* · Clustered Bar Charts
* · Overlapped Bar Charts
* · Lollipop Charts
* · Dumbbell Charts
* · Diverging Bars

**Standard Bar Charts**

Standard Bar Charts (SBCs) **compare only one numerical variable** **per item or category**. They try to answer the question: “How many are in each category”? Remember that categories or items refer to qualitative elements such as countries, cities, surnames, companies, brands, years, months, dates, data science methods, and so on.

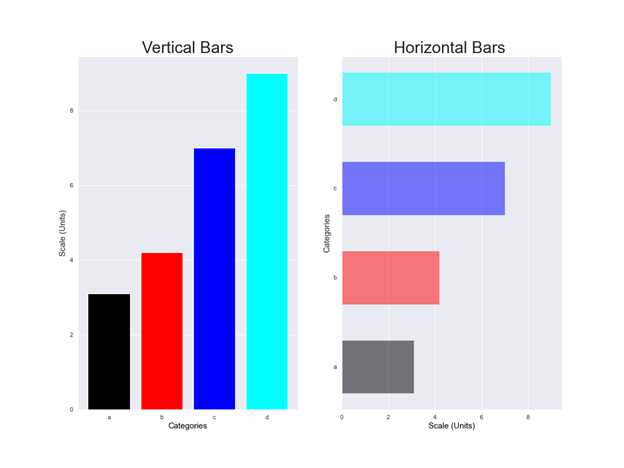


Fig. 1: Schematic layout of vertical and horizontal SBCs. Made by the author with Matplotlib.

There are vertical bar charts (column charts) and horizontal bar charts. Column charts compare items through the height of vertical rectangular bars. Horizontal bar charts compare quantities through the length of horizontal rectangular bars. The final value of the rectangle (length or height of each bar) is proportional to the numerical value that is intended to be compared. Each bar represents a single item and some space is habitually left between them.

Audiences are familiar with these types of charts so that they can focus on the message without wasting time comprehending the diagram. More details can be found in my [previous article](https://medium.com/nightingale/bar-graphs-why-how-8c031c224c9f).

**Clustered Bar Charts**

Clustered Bar Charts (CBCs) show numerical information about the relative proportion that exists between a main item or category and its subgroups that belongs to a second categorical variable. A variable is called categorical if its observations can be assigned into non-overlapping categories. Usually, they can take one of a limited (usually fixed) number of possible values.

Similar to SBCs, they can be oriented horizontally or vertically. Each principal item is divided into a **cluster of bars** representing subgroups of the second categorical variable. The quantity of each subcategory is shown by the height or length of some rectangular bars that are located side by side forming such cluster, with gaps between clusters slightly wider than a single standard bar.

CBCs are used for storytelling of Comparison and Proportion, but with an emphasis on Composition (**Part of a Whole Analysis**). Consequently, CBCs are particularly efficient when a whole is divided into **multiple parts**. They enable to make **comparisons** **across subgroups** (Stacked Bar Graphs make comparisons within subgroups).

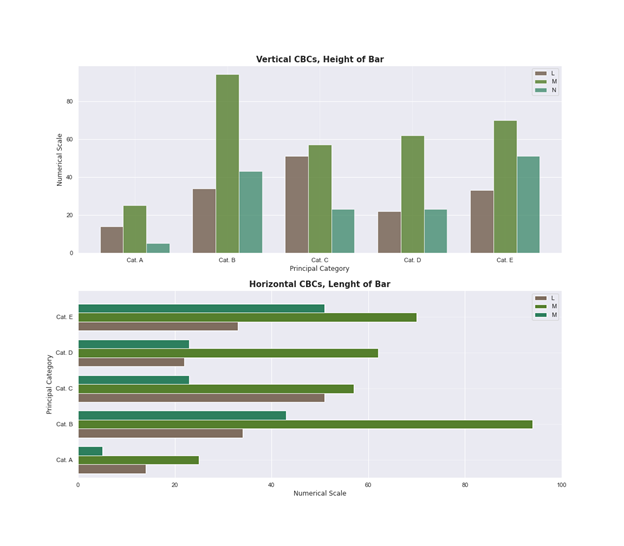


Fig. 2: Schematic layout of vertical and horizontal CBCs. Made by the author with Matplotlib.

The following figure displays information about a fictitious company related to sales, expenses, and profits for the 2016–2019 period. It is a vertically oriented CBC with years as the principal category. Sales, expenses, and profit are yearly represented as a cluster of bars. The chart shows that in 2018, although expenses increased significantly, profits were only reduced by a small amount.

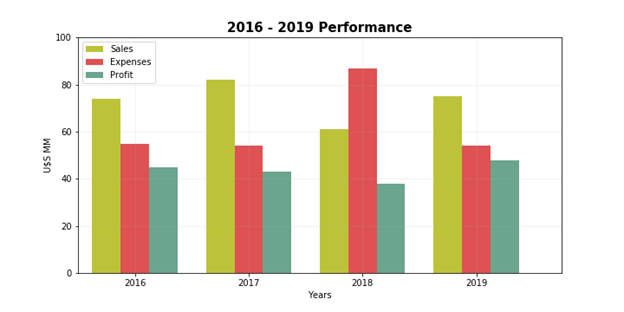


Fig. 3: Economic performance of a fictitious company during the 2016–2019 period. The figure was developed by the author with Matplotlib.

Audiences are also familiar with these types of charts so that they can focus on the message without wasting time in comprehending the diagram. More details can be found in my [previous article](https://medium.com/towards-data-science/clustered-overlapped-bar-charts-94f1db93778e).

**Overlapped Bar Charts**

Overlapped Bar Charts (OVCs) are used when we want **to compare two numerical variables per item** or category in a single diagram. Of course, both numerical variables must be sufficiently related to justify a comparison.

The conceptual idea is to contrast the numerical values of two variables by means of their overlapping. Such overlapping allows us to tell the story with greater expositional power.

Similar to SBCs and CBCs, the classic layout is with two axis and rectangular bars that can be oriented horizontally or vertically. One axis shows categories, the other axis shows numerical values related to the variables to be compared. Of course, both numerical variables must share the same numerical scale. The width of the bars is different for each numerical variable with the smaller going forward for clarity of reading, although this situation may not always be possible for all charts.

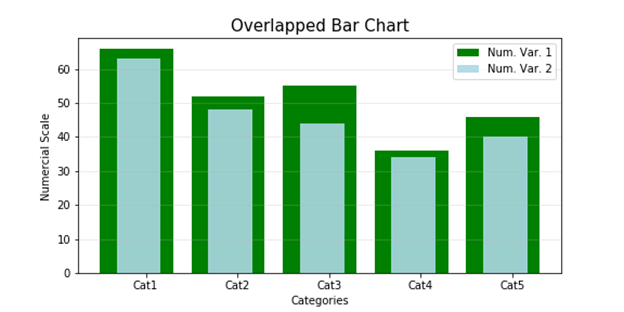


Fig. 4: An overlapped bar chart developed by the author with Matplotlib.

More than two numerical variables can be compared with a **partially overlapped bar chart.** In such type of charts, bars representing different numerical variables are partially hidden by other rectangles located in front of them. From a conceptual point of view, a partially overlapped BC is like a CBC when the rectangles representing the different subgroups begin to overlap instead of being located side by side. Great care must be taken to avoid confusing the audience. In this sense, it is only recommended to use them to compare up to three different numerical variables.

More details can be found in my [previous article.](https://medium.com/towards-data-science/clustered-overlapped-bar-charts-94f1db93778e)

**Lollipop Charts**

Conceptually similar to SBCs, Lollipop Charts (LCs) are used to make **comparisons** between different items or categories. They compare only one numerical variable per item. They differ in that the LCs replace the rectangular bar with a line with a dot at the end. The corresponding numerical value is indicated by the **position of the dot** at the end of such line. So, the locations of the dots at the end of those lines are equivalent to the height or length of vertical or horizontal standard bars.

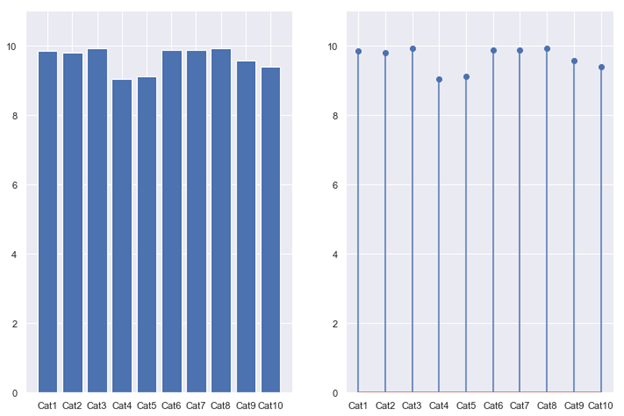


Fig. 5: Comparison between a standard bar chart and a lollipop chart. Made by the author with Matplotlib.

The classic layout is with two axis and very thin lines that can be oriented vertically or horizontally. One axis indicates categories, and the other axis has a numerical scale (preferably with units) related to the items to be compared.

It is advisable to use LCs as an alternative to SBCs when there is a need to display a **large number of similar values.** In this way, we avoid displaying a cluttered chart and we also prevent our audience from experiencing an annoying optical effect called Moiré pattern.

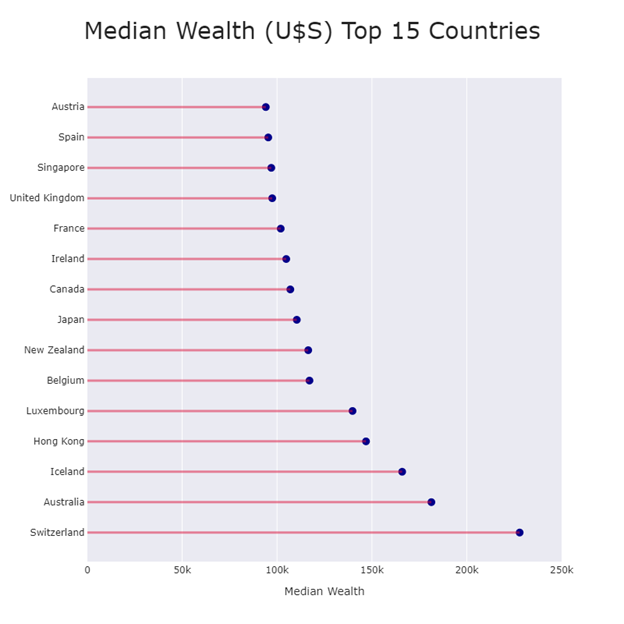


Fig.6: A Lollipop chart made by the author with Plotly.

**Dumbbell Charts**

While Dumbbell Charts are quite similar to LCs, their main objective is to indicate change between two data points. In that sense, they are habitually used **to make comparisons between two categories.**

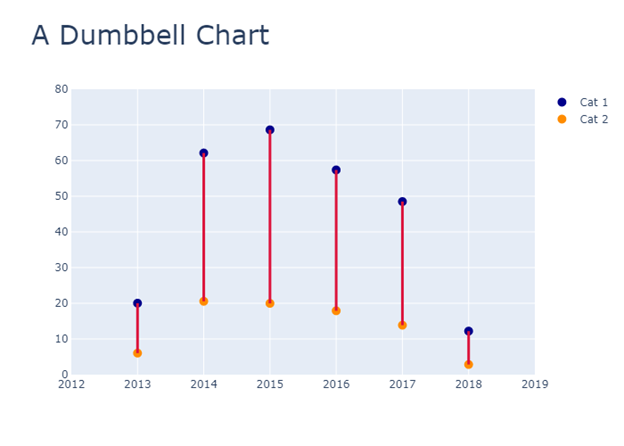


Fig. 7: A Dumbbell chart made by the author with Plotly.

We recommend using them for comparing ranges, spread, changes, and differences between two numerical variables or distances between points.

**Diverging Bar Charts**

A diverging bar chart (DBC) consists basically of two horizontal rectangles (bars) aligned so that one of the rectangles runs from right to left, the other from left to right, and both start from a common vertical baseline, usually located in the center of the chart. As already indicated, the length of each rectangle (bar) is proportional to the numerical value it is intended to display. Each bar represents an item or categorical variable and some space must be left between them. The best encoding for the DBCs is when two numerical alternatives are to be compared.

The Butterfly Chart is a particular type of DBC. Usually, they leave some space between the bars where the names of the variables being compared are placed.

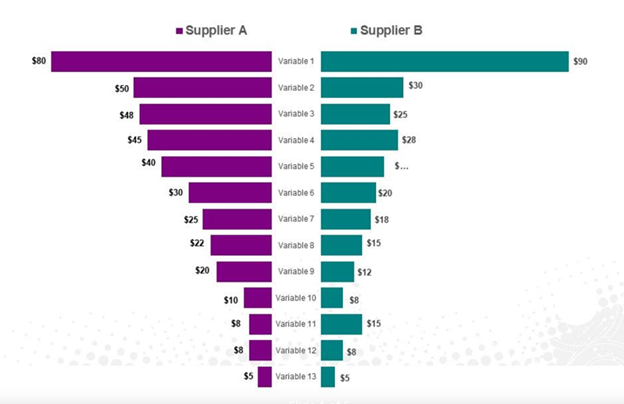


Fig.8: A butterfly chart. Source: #1.

Another set of charts can also be used to compare items: radar charts, Mekko charts, and Marimekko Charts. The problem is that they are very difficult to interpret because they are based on audiences’ ability to decode numerical information by comparing angles or areas. Therefore, it is always preferable to select some of the charts previously indicated. More details can be found in my [previous article.](https://medium.com/towards-data-science/mekko-charts-f38311c576e2)

Data visualization is the most powerful tool to tell the story behind our data. But our audience can be confused if we do not choose correctly the visualization technique best suited to the message we want to show.

This is the aim of this series of articles: to indicate which charts and diagrams are the most appropriate for a particular type of message.

In this article, we indicated six different charts [Standard Bar Charts, Clustered Bar Charts, Overlapped Bar Charts, Lollipop Charts, Dumbbell Charts, Diverging Bars] **that aim to compare items.** We indicated similarities and differences between them.

It is imperative that data scientists and data analysts receive training related to the different types of messages that audiences require to perform their tasks. Associated with the above, is the selection of the most suitable chart to tell the story behind our numbers.

Stay tuned for the upcoming articles.