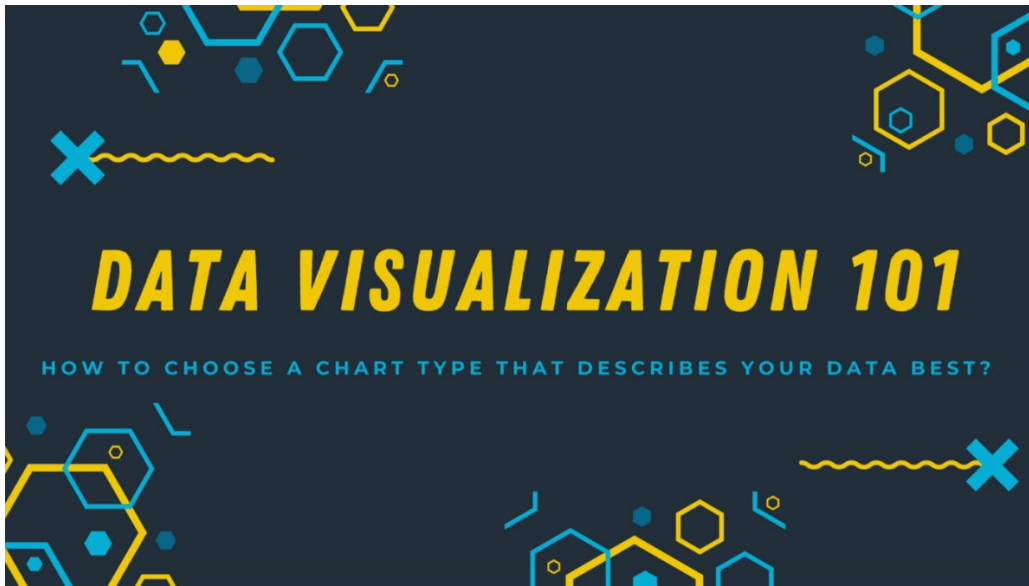


Data Visualization 101: How to Choose a Chart Type

How to choose a chart type that describes your data best??



When working on any data science project, one of the essential steps to explore and interpret your results is to visualize your data. At the beginning of the project, visualizing your data helps you understand it better, find patterns and trends.

At the end of the project, after you've done your analysis and applied different machine learning models, data visualization will help you communicate your results more efficiently.

Humans are visual creatures by nature; things make sense to us when it's represented in an easy to understand visualization. It's way easier to interpret a bar chart than it is to look at massive amounts of numbers in a spreadsheet.

Efficient data visualization can make or break your project. If you put tons of effort into analyzing and modeling your data, but you ended up using

the wrong chart type to present your results, your audience will not grasp the effort you put in or how to use these results.

There are many chart types, so many, the process of choosing the correct one can be overwhelming and confusing. This article will — hopefully — give you a simple and straightforward approach to selecting the best chart type that represents your data perfectly and communicate it most efficiently.

How to start?

Before you start looking at chart types, you need to ask yourself 5 critical questions about your data. These questions will help you understand your data better and hence, choose the perfect chart type to represent it.

№1. What's the story your data is trying to deliver?

Data is just a story told in numbers.

So, the first thing you need to know about your data is, what story is it trying to deliver? Why was this data collected, and how?

Is your data collected to find trends? To compare different options? Is it showing some distribution? Or is used to observe the relationship between different value sets?

Understanding the origin story of your data and knowing what it's trying to deliver will make choosing a chart type a much easier task for you.

№2. Who will you present your results to?

Once you figured out the story behind your data, next, you need to know who you will be presenting your results for. If you're analyzing stock market trends and you will present your findings to some businessmen, you might use a different chart type than if you were representing your finding for people getting started with the stock market.

The whole purpose of using data visualization is to make data communication more efficient.

For that reason, you need to know your audience so you can choose the best chart type to use when representing your data to them.

№3. How big is your data?

The size of your data will significantly affect the type of chart you will use. Some types of charts are not meant to be used with massive datasets, while others are perfect for big data.

For example, piecharts work best with a small number of datasets; however, if you're using a significant amount of datasets, using a scatter plot will make more sense.

You need to select a chart type that fits the size of your data best and represents it clearly without cluttering.

№4. What is your data type?

There are several types of data, describe, continuous, qualitative, or categorical. You can use the kind of data to eliminate some chart types. For example, if you have continuous data, a bar chart may not be the best choice; you may need to go with a line chart instead.

Similarly, if you have categorical data, then using a bar chart or a pie chart may be a good idea. You probably will not want to use a line chart with categorical data, because by definition, you can't have continuous categories. There has to be a discrete finite amount of categories.

№5. How do the different elements of your data relate to each other?

Finally, you need to ask yourself how do the different elements of your data relate. Is your data order based on some factor — time, size, type?

Doesn't represent a ranking based on some variable? Or a correlation between different variables?

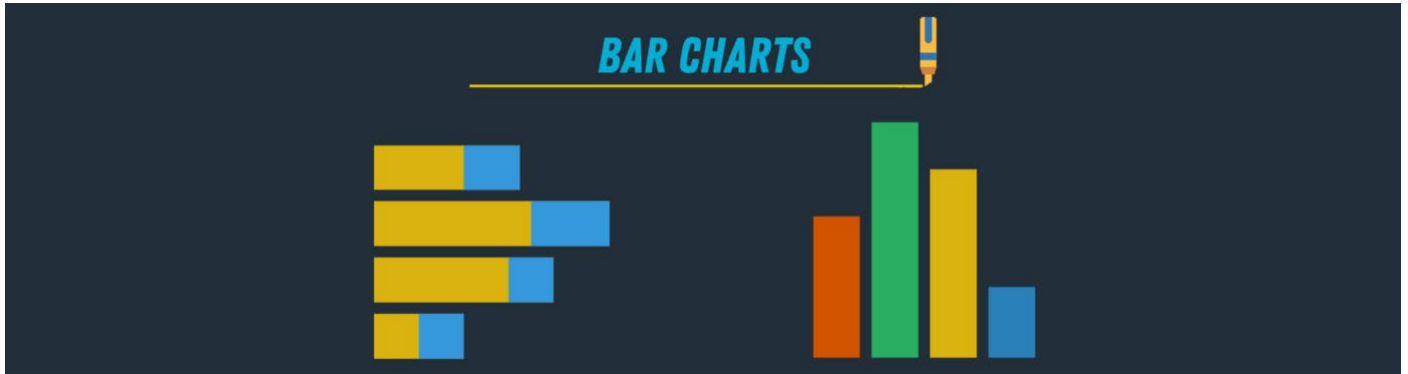
Is your data a time-series — data that changes over time? Or is it more of a distribution?

The relationship between the values within your dataset may decide on what chart type to use a bit more straightforward.

The top 7 used chart types

There are more than 40 types of charts out there; some are more commonly used than others because they are easier to build and interpret. Let's talk about the top 7 used charts type and when to use each of them.

Bar Chart



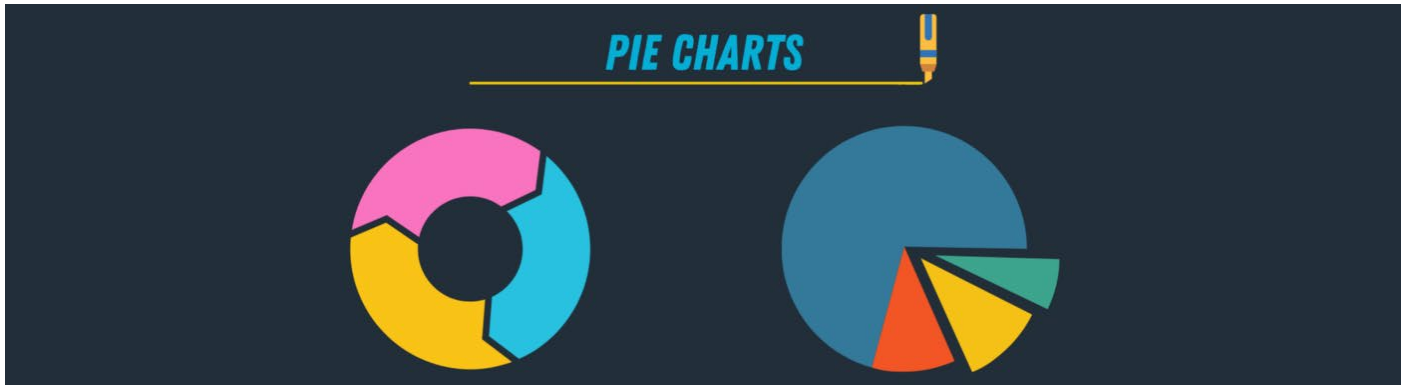
When to use:

1. Comparing parts of a bigger set of data, highlighting different categories, or showing change over time.
2. Have long categories label — it offers more space.
3. If you want to illustrate both positive and negative values in the dataset.

When to avoid:

1. If you're using multiple data points.
2. If you have many categories, avoid overloading your graph. Your graph shouldn't have more than 10 bars.

Pie Chart



When to use:

1. When you show relative proportions and percentages of a whole dataset.
2. Best used with small datasets — also applies to donut charts.
3. When comparing the effect of ONE factor on different categories.
4. If you have up to 6 categories.
5. When your data is nominal and not ordinal.

When to avoid:

1. If you have a big dataset.
2. If you want to make a precise or absolute comparison between values.

Line Chart



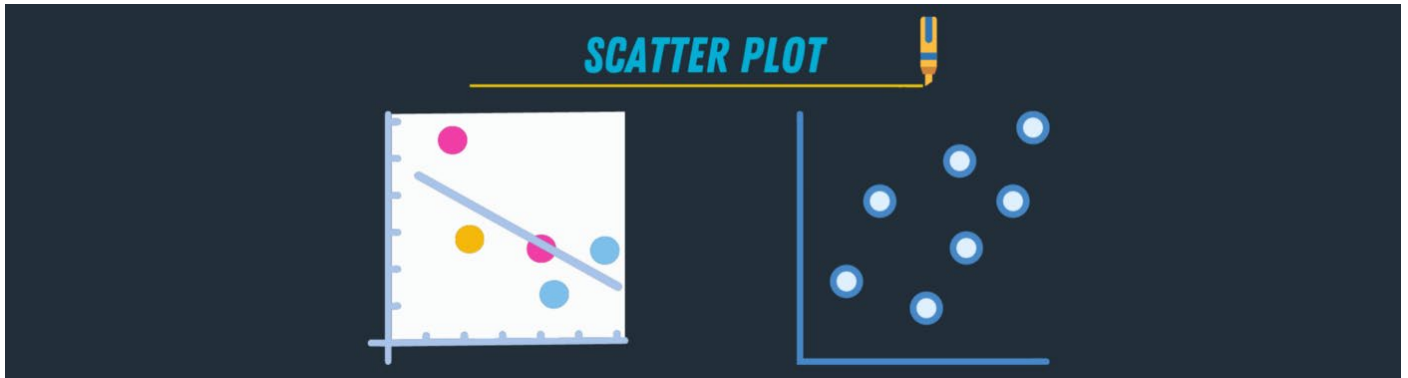
When to use:

1. If you have a continuous dataset that changes over time.
2. If your dataset is too big for a bar chart.
3. If you want to display multiple series for the same timeline.
4. If you want to visualize trends instead of exact values.

When to avoid:

1. Line charts work better with bigger datasets, so, if you have a small one, use a bar chart instead.

Scatter Plot



When to use:

1. To show correlation and clustering in big datasets.
2. If your dataset contains points that have a pair of values.
3. If the order of points in the dataset is not essential.

When to avoid:

1. If you have a small dataset.
2. If the values in your dataset are not correlated.

Area Chart



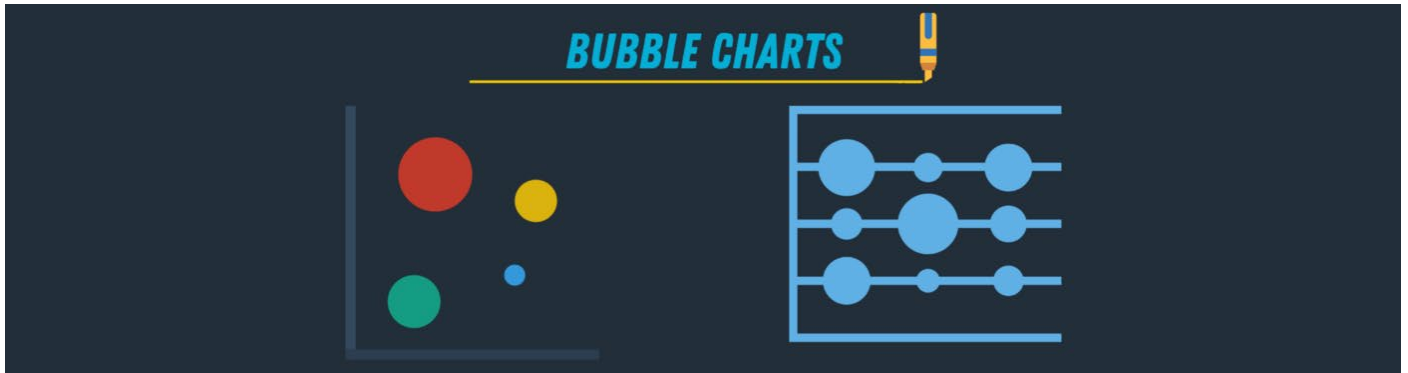
When to use:

1. If you want to show part-to-whole relations.
2. If you want to portray the volume of your data and not just the relation to time.

When to avoid:

1. It can't be used with discrete data.

Bubble Chart



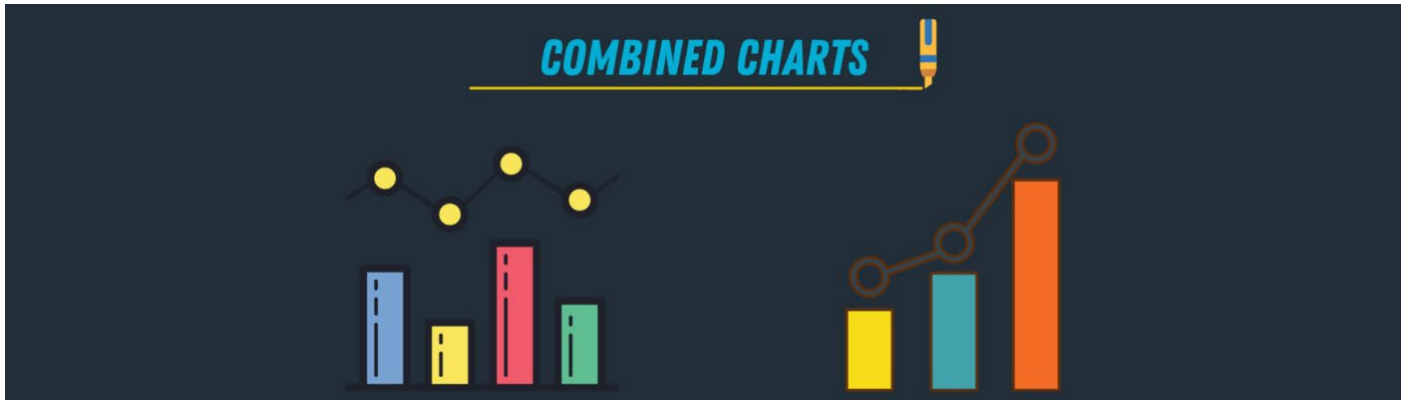
When to use:

1. If you want to compare independent values.
2. If you want to show distribution or relation.

When to avoid:

1. If you have a small dataset.

Combined Chart



When to use:

1. If you want to compare values with different measurements.
2. If the values are different in range.

When to avoid:

1. If you want to display more than 2~3 types of graphs. In that case, it's better to have separate graphs to make it easier to read and understand.

Chart selection tips

Whenever you decide to create some data visualization, use these best practices to make it more straightforward and effective.

1. If you have categorical data, use a bar chart if you have more than 5 categories or a pie chart otherwise.
2. If you have nominal data, use bar charts or histograms if your data is discrete, or line/ area charts if it is continuous.
3. If you want to show the relationship between values in your dataset, use a scatter plot, bubble chart, or line charts.
4. If you want to compare values, use a pie chart — for relative comparison — or bar charts — for precise comparison.
5. If you want to compare volumes, use an area chart or a bubble chart.
6. If you want to show trends and patterns in your data, use a line chart, bar chart, or scatter plot.

Conclusion

Before you choose what chart type to use, you need to get to know your data better, the story behind it, and your target audience/media.

Whenever you try to create a visualization, chose simple colors and fonts.

Always aim for simple visualization than complex ones. The goal of visualizing data is to make it easier to understand and read. So, avoid overloading and cluttering your graphs. Having multiple simple graphs is always better than one elaborate graph.