

Data Visualization

What is data visualization?

Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. The main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets. The term is often used interchangeably with others, including information graphics, information visualization and statistical graphics.

Data visualization is one of the steps of the data science process, which states that after data has been collected, processed and modeled, it must be visualized for conclusions to be made. Data visualization is also an element of the broader data presentation architecture (DPA) discipline, which aims to identify, locate, manipulate, format and deliver data in the most efficient way possible.

Data visualization is important for almost every career. It can be used by teachers to display student test results, by computer scientists exploring advancements in artificial intelligence (AI) or by executives looking to share information with stakeholders. It also plays an important role in big data projects. As businesses accumulated massive collections of data during the early years of the big data trend, they needed a way to get an overview of their data quickly and easily. Visualization tools were a natural fit.

Advantages of Data Visualization

- 1 Improved agreement.
- 2 A better way.
- 3 Simple data sharing.
- 4 Thorough inquiry.
- 5 Transaction's inquiry.

Disadvantages of Data Visualization

- 1 Provides an evaluation rather than exactness.
- 2 Unidirectional.
- 3 Lack of assistance.
- 4 An insufficient plan.

Types of data visualizations

- **Tables:** This consists of rows and columns used to compare variables. Tables can show a great deal of information in a structured way, but they can also overwhelm users that are simply looking for high-level trends.
- **Pie charts and stacked bar charts:** These graphs are divided into sections that represent parts of a whole. They provide a simple way to organize data and compare the size of each component to one other.
- **Line charts and area charts:** These visuals show change in one or more quantities by plotting a series of data points over time and are frequently used within predictive analytics. Line graphs utilize lines to demonstrate these changes while area charts connect data points with line segments, stacking variables on top of one another and using color to distinguish between variables.
- **Histograms:** This graph plots a distribution of numbers using a bar chart (with no spaces between the bars), representing the quantity of data that falls within a particular range. This visual makes it easy for an end user to identify outliers within a given dataset.
- **Scatter plots:** These visuals are beneficial in revealing the relationship between two variables, and they are commonly used within regression data analysis. However, these can sometimes be confused with bubble charts, which are used to visualize three variables via the x-axis, the y-axis, and the size of the bubble.
- **Heat maps:** These graphical representation displays are helpful in visualizing behavioral data by location. This can be a location on a map, or even a webpage.
- **Tree maps,** which display hierarchical data as a set of nested shapes, typically rectangles. Treemaps are great for comparing the proportions between categories via their area size

Tips for data Visualization:

1. Data visualizations should be audience-specific with a clear requirement
2. Choose the right visualization for your data
3. Keep your visualizations simple
4. Label your data visualizations
5. Understand the importance of text in charts
6. Use colors effectively in data visualizations
7. Avoid deceiving with your visualizations
8. Make interpretable data visualizations

1. Data visualization should be audience-specific with a clear requirement

Amongst the data visualization tips, this is the first one. While creating data visualizations, it is important to know the requirement of the chart and the audience it is for. These two things alone can take your visualization from zero to hero. This makes sure that you not only create a visualization with a strategic purpose that answers a specific question but also one that can be easily understood by the audience.

For example, if your audience doesn't have a background in science, then don't create a visualization that is filled with scientific information. Similarly, bombarding your chart with multiple trends will most likely divide the attention of the viewer and defeats the purpose of the visualization.

Know the requirement of the visualization. This allows you to create a chart that conveys a message in a clear and crystal manner. Also, it makes sure that you aren't overloading your chart with unnecessary information that might confuse the audience.

2. Choose the right data visualization for your data

Amongst the data visualization tips, this one is of utmost importance. There is a myriad of visualization graphs out there. But choosing the right

one is important so as to effectively highlight the key trend in the data. Also, choosing the right graph for your visualization will make sure the message is easy to grasp and the viewers are attracted to your work. Each graph has a specific purpose and one should know where to use which graph.

Bar graphs are one of the most popular types of data visualizations. They offer a great amount of information in a quick glance. They are best to compare a few values within the same category. For example, comparing the sales of two different products over the years.

Line plots are useful for visualizing the trend in a numerical value over a continuous time interval. They effectively capture the trends and patterns in data and can be used to compare multiple values. An example of such a data visualization would be to show the trend in the monthly income of a company over the last few months.

Scatter plots are useful for showing the relationship between two variables. Any correlation between variables or outliers in the data can be easily spotted using scatter plots. For example, it can be used to compare how the price of a house varies with the size of the living room.

Pie charts are suitable to show the proportional distribution of items within the same category. But they should be used prudently otherwise they do more harm than good. For example, the percentage of android users to iOS users in a country.

Histograms show the distribution of numeric data through a continuous interval by segmenting data into different bins. They are great for showing the distribution of data. For example, visualizing the number of orders for a product over the years.

And so on. Also, don't be afraid to combine more than one type of graph in your visualizations. Sometimes it offers a chance for the viewer to explore the data in detail.

3. Keep your visualizations simple

It is very easy to put up too much information in a visualization. But harder to get rid of the unnecessary information. A minimalist visualization that is devoid of distractions and unnecessary patterns is likely to convey the message to the viewer more effectively.

Therefore, it is best to pay close attention to this data visualization technique and only add those elements to your visualization that provide a value-for-money and simplify the chart for the viewer.

4. Label your data visualizations

An important data visualization technique is to label your visualization. This better conveys what the visuals are trying to say. They are an easy miss when creating the visualization, so make sure you double-check for labeling before rolling out your visualization.

Labels should be legible. If it is not clear, it is of no use. Therefore, make sure the labels are easy to read and comprehensible. Give a title to the graph. Viewers can easily get instant gist of what the graph is about when you give a suitable title to your graph. Use a legend wisely. A legend makes it easier to spot the difference between the various lines in the graph. But when using line charts, try to label directions. This makes it easier to identify lines.

Label your axes. Sometimes it might not be clear from the title what the axes represent. Therefore, you might want to label your axes at times.

Pay attention to the labeling on the axes. Sometimes, you don't need to label all the ticks on the axes. You can instead label them at intervals if they still convey the right message.

5. Understand the importance of text in charts

Data visualization is not just about numbers. The text provides an important context that conveys the right message to the viewer. Headings, sub-headings, and annotations that you put alongside the graphs, explain what is being presented in the visualization. But reiterating the same message in every text and unnecessarily putting too much text can backfire. It can end

up doing more harm than good. Therefore, it is best to use text in moderation.

1. Try using simple phrases wherever possible. The aim is to allow the visualization to speak for itself.
2. Keep only those annotations that provide relevant information. Putting up annotations for every data point will distract the viewer and unnecessarily clutter the visual.
3. You might need to use bold or italic text to highlight important parts of the graph, but try not to use them excessively otherwise there will be no difference between regular and emphasized text.
4. Avoid text reiterating the same message. For example, the heading and subheading repeating the same message might not be the prudent thing to do.
5. Avoid using distracting fonts that are hard to read. The viewer should be able to grasp the message in the graph instantly without much work.

6. Use colors effectively in data visualizations

Everyone knows the power of colors and the amount of impact it can have on the viewer. It is one of the most important data visualization tricks that you can employ for your visualization. It can provide the right amount of zest that your visualization needs to entice the viewers. But improper use of colors can end up misleading the viewer. Therefore, the data visualization technique requires close attention.

7. Avoid deceiving with your visualizations

While we are trying so much to create a stunning visualization, it can very easily deceive the viewer. And sometimes we won't even know it while deceiving the viewer. Small things like cherry-picking data, omitting baseline, information overload, etc. can all lead to deception. Therefore, one should avoid such naive mistakes while creating visualizations.

Stock market visualizations are a classic example where deception is very common. If you don't display the complete picture, you can get a false idea of how a company is doing. The visualization on the left gives the

performance of the company in the past month. It looks like the company isn't doing so well. But, if we look at the stock prices for the past 6 months, we can see that it actually went up and the company is doing relatively well!

8. Make interpretable data visualizations

The last of the data visualization tips is that the interpretability of the visualization matters more than its visual appeal. All the points we have covered so far should make the visualization more interpretable. Visuals images, patterns, colors, etc. are only good if they don't distort the message to the viewer. In the end, even if a simple line graph is able to deliver the message across to the viewer, then you don't really need to put fancy logos or images in your visualizations!

A better way to represent such data would be to simply use a bar chart which would have accentuated the rankings of the different languages.

Open source visualization tools

Access to data visualization tools has never been easier. Open source libraries, such as D3.js, provide a way for analysts to present data in an interactive way, allowing them to engage a broader audience with new data. Some of the most popular open source visualization libraries include:

- **D3.js:** It is a front-end JavaScript library for producing dynamic, interactive data visualizations in web browsers. D3.js (link resides outside IBM) uses HTML, CSS, and SVG to create visual representations of data that can be viewed on any browser. It also provides features for interactions and animations.
- **ECharts:** A powerful charting and visualization library that offers an easy way to add intuitive, interactive, and highly customizable charts to products, research papers, presentations, etc. Echarts (link resides outside IBM) is based in JavaScript and ZRender, a lightweight canvas library.

- **Vega:** Vega (link resides outside IBM) defines itself as “visualization grammar,” providing support to customize visualizations across large datasets which are accessible from the web.
- **deck.gl:** It is part of Uber's open source visualization framework suite. deck.gl (link resides outside IBM) is a framework, which is used for exploratory data analysis on big data. It helps build high-performance GPU-powered visualization on the web.

Data visualization best practices

With so many data visualization tools readily available, there has also been a rise in ineffective information visualization. Visual communication should be simple and deliberate to ensure that your data visualization helps your target audience arrive at your intended insight or conclusion. The following best practices can help ensure your data visualization is useful and clear:

Set the context: It's important to provide general background information to ground the audience around why this particular data point is important. For example, if e-mail open rates were underperforming, we may want to illustrate how a company's open rate compares to the overall industry, demonstrating that the company has a problem within this marketing channel. To drive an action, the audience needs to understand how current performance compares to something tangible, like a goal, benchmark, or other key performance indicators (KPIs).

Know your audience(s): Think about who your visualization is designed for and then make sure your data visualization fits their needs. What is that person trying to accomplish? What kind of questions do they care about? Does your visualization address their concerns? You'll want the data that you provide to motivate people to act within their scope of their role. If you're unsure if the visualization is clear, present it to one or two people within your target audience to get feedback, allowing you to make additional edits prior to a large presentation.

Choose an effective visual: Specific visuals are designed for specific types of datasets. For instance, scatter plots display the relationship between two variables well, while line graphs display time series data well. Ensure that the visual actually assists the audience in understanding your main takeaway. Misalignment of charts

and data can result in the opposite, confusing your audience further versus providing clarity.

Keep it simple: Data visualization tools can make it easy to add all sorts of information to your visual. However, just because you can, it doesn't mean that you should! In data visualization, you want to be very deliberate about the additional information that you add to focus user attention. For example, do you need data labels on every bar in your bar chart? Perhaps you only need one or two to help illustrate your point. Do you need a variety of colors to communicate your idea? Are you using colors that are accessible to a wide range of audiences (e.g. accounting for color blind audiences)? Design your data visualization for maximum impact by eliminating information that may distract your target audience.