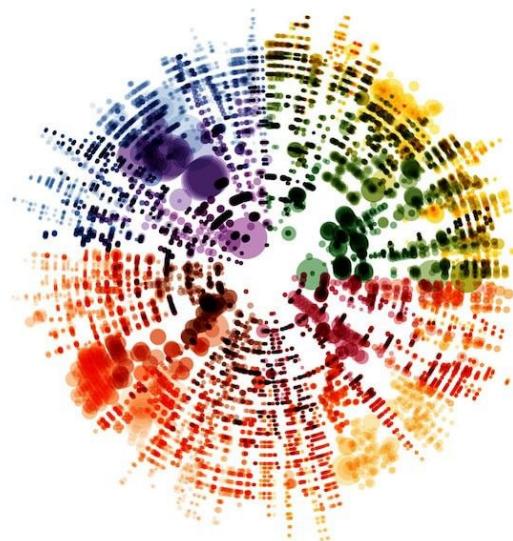


Tips for Data Visualization

Your data is only as good as your ability to understand and communicate it, which is why choosing the right visualization/chart is essential. If you are unable to present the data effectively, key insights may be lost. Information can be visualized in several different ways, each of which is better at providing a specific insight.



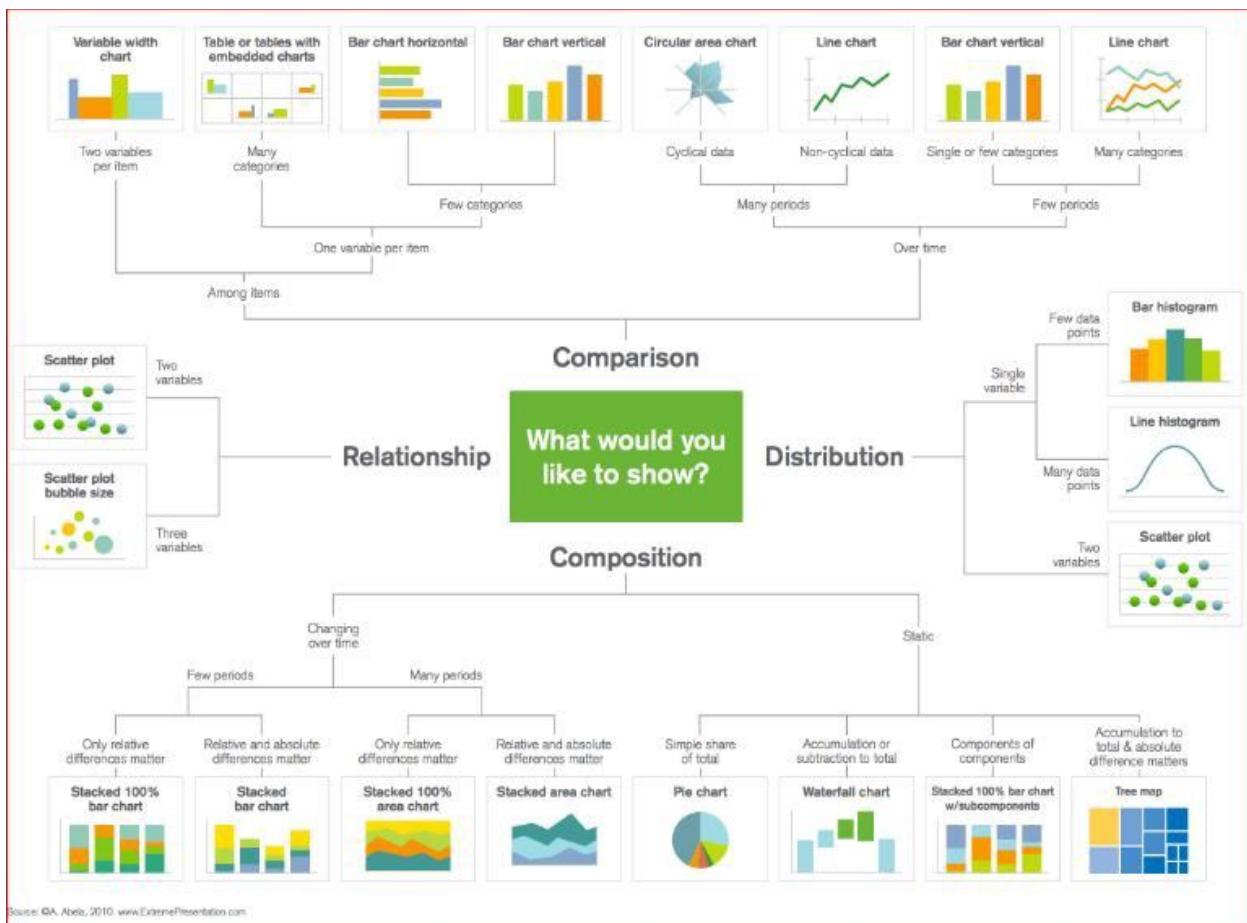
When working with data, start with understanding the story you want to tell or relationship you are trying to highlight. This will aid in selecting the right visualization that will deliver your message.

Before starting to visualize data, you should understand the **common types of data:**

1. Quantitative — can be counted or measured; numerical
2. Categorical — can be sorted according to groups or categories
3. Discrete — has a finite number of possible values
4. Continuous — measured and has a value within a range

Data Visualization is a big part of a data scientist's jobs. In the early stages of a project, you'll often be doing an Exploratory Data Analysis (EDA) to gain some insights into your data. Creating visualizations really helps make things clearer and easier to understand, especially with larger, high dimensional datasets. Towards the end of your project, it's important to be able to present your final results in a clear, concise, and compelling manner that your audience, whom are often non-technical clients, can understand.

Here's a great chart for selecting the right visualization for the job!



Understanding **data relationships** is important to the data visualization process. Common data relationships are listed below:

- 👉 Time Series — tracks changes in values of a metric over time
 - 👉 Nominal Comparison — compares quantitative values of subcategories
 - 👉 Correlation — data with two or more variables that may have a positive or negative correlation to each other
 - 👉 Part-to-Whole-Relationships — subset of data compared to the larger whole
 - 👉 Distribution — shows data distribution around a central value
 - 👉 Deviation — shows how data points relate to each other, to see how far any given point is from the mean
 - 👉 Ranking — shows how two or more values compare to each other in relative magnitude
-

There are various packages available in Python for visualization purpose. Some of them are: **matplotlib**, **seaborn**, **bokeh**, **plotly**, **ggplot**, **pygal**, **gleam**, **leather** etc. Each package has its own utility and which package to use depends mainly on the type of dataset, problem to solve and what you want to infer from it.

On the basis of the number of variables you want to see relationship in between, plotting can be classified as:

1. Univariate plotting
 - Bar charts
 - Area charts
 - Line Charts
 - Histograms

2. Bivariate plotting

- Scatter Plot
- Hexplot
- Stacked Plots
- Bivariate Linear charts
- Pair Plot

3. Multivariate plotting

- Multivariate scatter plot
- Grouped Boxplot
- Heat map
- Parallel Coordinates
- Swarm plot

Other tips:

- 👉 Use icons to enhance understanding and reduce extra labels/legends
- 👉 Order data sets using a logical hierarchy
- 👉 Visualize data in a way that makes it simple for the audience to compare values
- 👉 Highlight important information using annotations or callouts
- 👉 Use one color to represent each category, consistently throughout the presentations/ dashboards
- 👉 Don't use 3D charts — it may skew understanding of the information
- 👉 Avoid using more than 5 colors in a view
- 👉 Don't use distracting background colors or those that have gradients