Charts: How to Choose the Right Chart for our Data?

Topics Covered:

- **Variety of charts & their properties**
- **Data Visualization Do's and Don'ts – A General Diagnosis**
- **Short summary of data visualizations types**
- **Summary for choosing appropriate chart based on properties of feature/column using seaborn**
- **Important link to explore**

## 1) Variety of Charts

a) Line Charts

- **A line chart reveals trends or changes over time.**
- **Line charts can be used to show associations within a continuous data set and can be applied to a wide variety of categories, including a daily number of customers to a shop or variations in gold prices.**
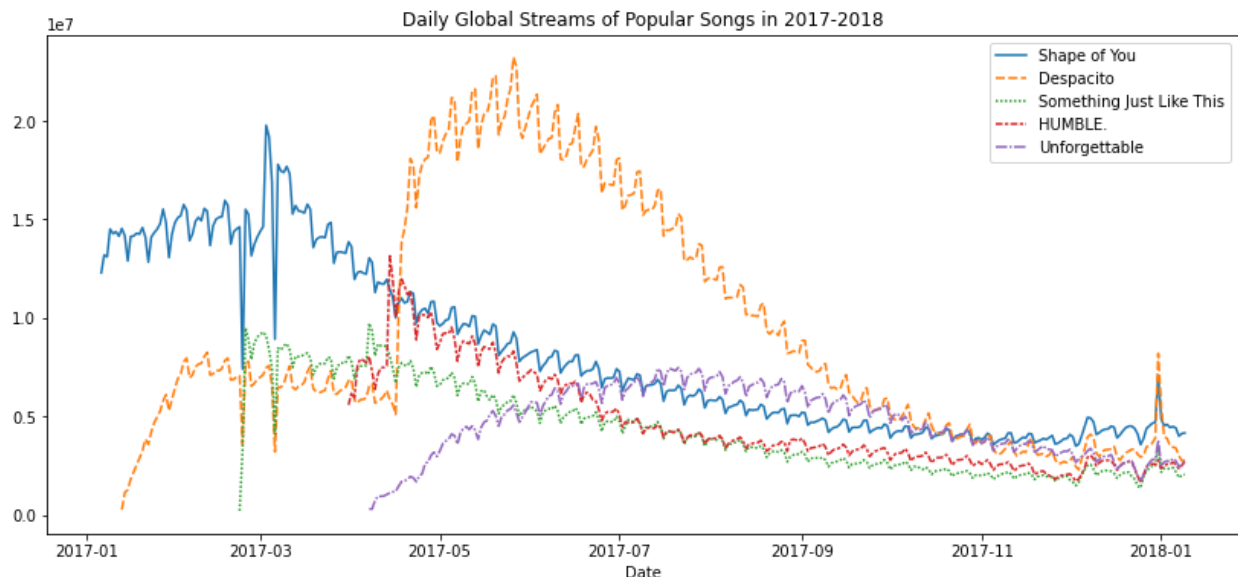

b) Pie Charts

- **Pie charts are used to show segments of an entity(whole).**
- **A pie chart illustrates numbers in percentages, and the total sum of all the divided sections equals 100 per cent.**

Immigration to Canada by Continent [1980 - 2013]
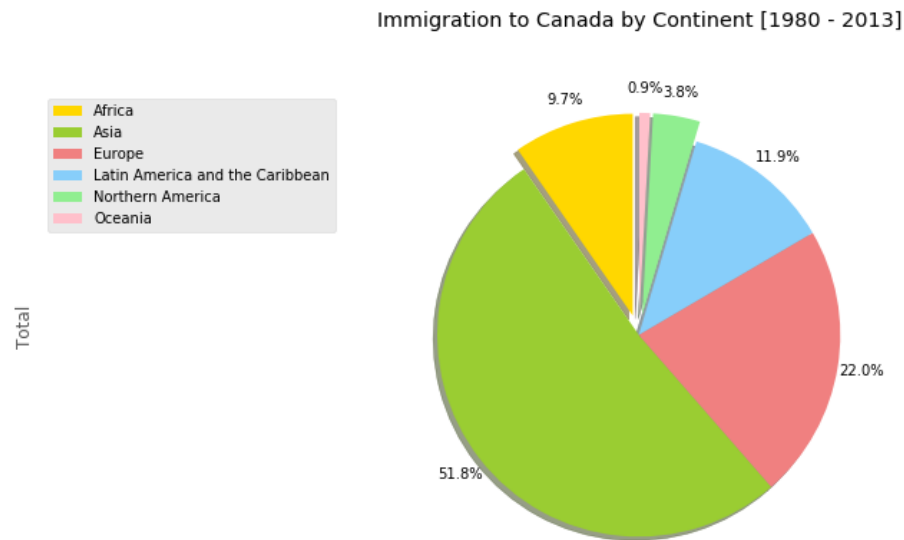
c) Column charts

- **The column chart is reasonably the most used chart type.**
- **A column chart is used to show a comparison among different items, or it can show a comparison of items over time.**
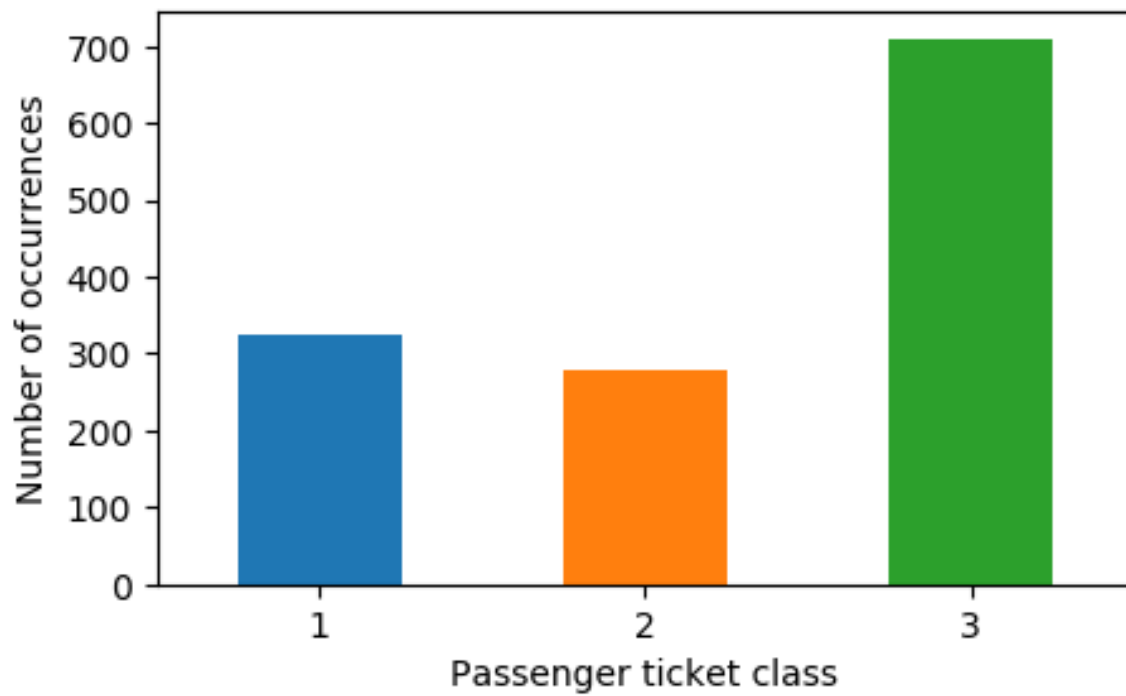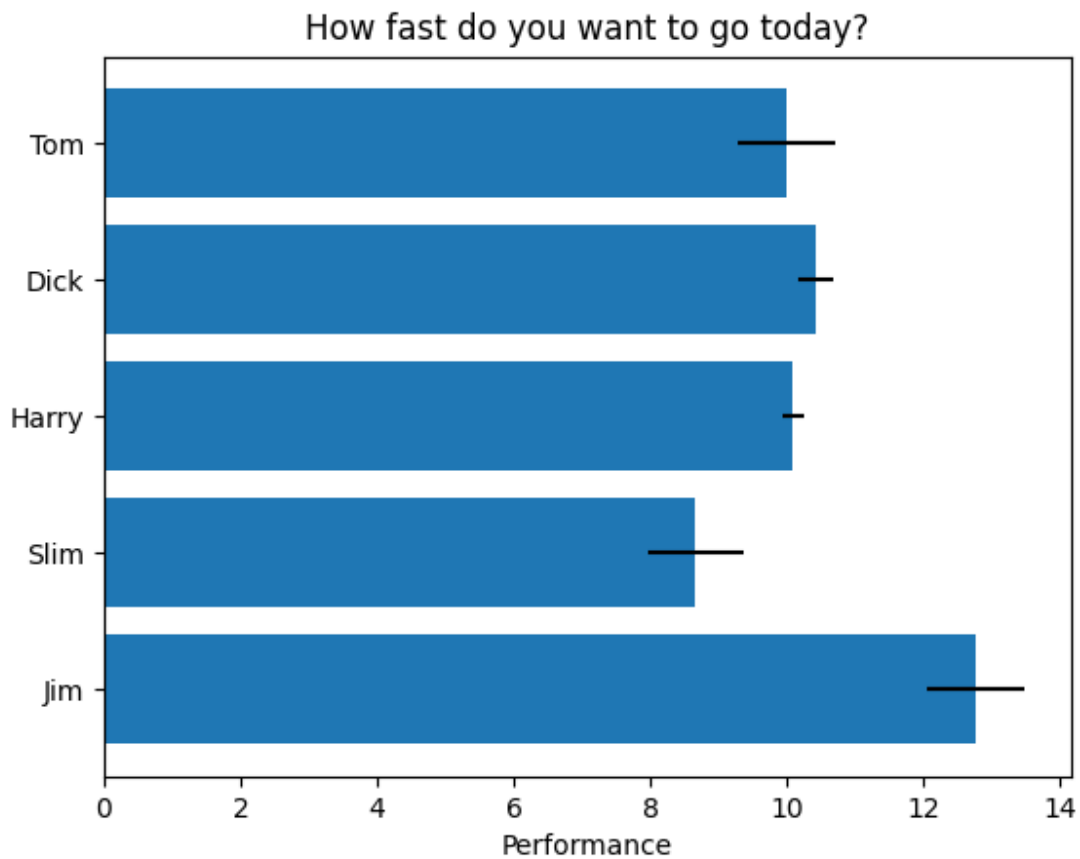
d) Stacked Column Charts

- Use stacked column charts to show a composition.
- Do not use too many composition items (not quite three or four) and confirm that
- composing parts are relatively similar in size.
- It can get messy very quickly.

Scores by group and gender
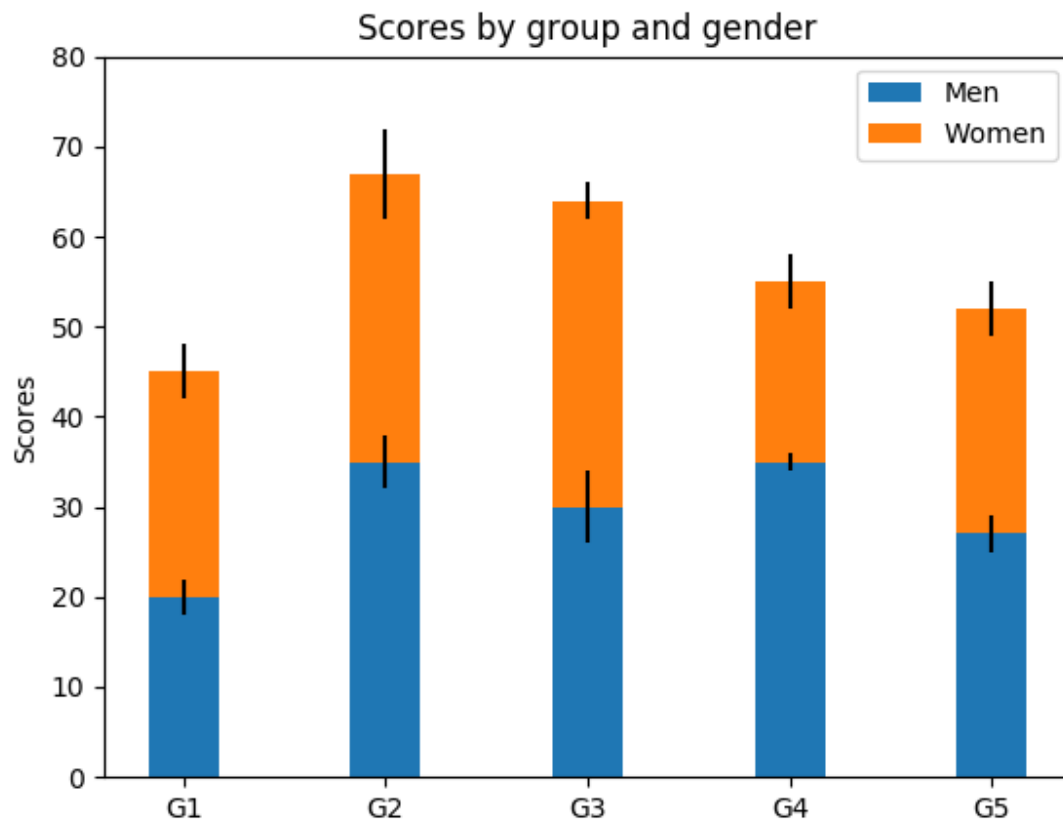
e) Bar Graphs

- **A bar chart, basically a horizontal column chart, should be used to avoid clutter when one data label is long or if you have got more than 10 items to match/differentiate.**
- **This type of visualization can also be used to display negative numbers.**

How fast do you want to go today?

A horizontal bar chart with categories Tom, Dick, Harry, Slim, and Jim on the y-axis and Performance on the x-axis (ranging from 0 to 14). Tom ≈ 10, Dick ≈ 10.5, Harry ≈ 10, Slim ≈ 8.5, Jim ≈ 12.7, each with error bars.

f) Area Charts

- **An area chart is essentially a line chart, but the space between the x-axis and the line is filled up with a colour or pattern.**
- **It is useful for showing part-to-whole associations, like showing individual sales reps' contribution to total sales for a year.**
- **It helps you analyze both overall & individual trend details.**
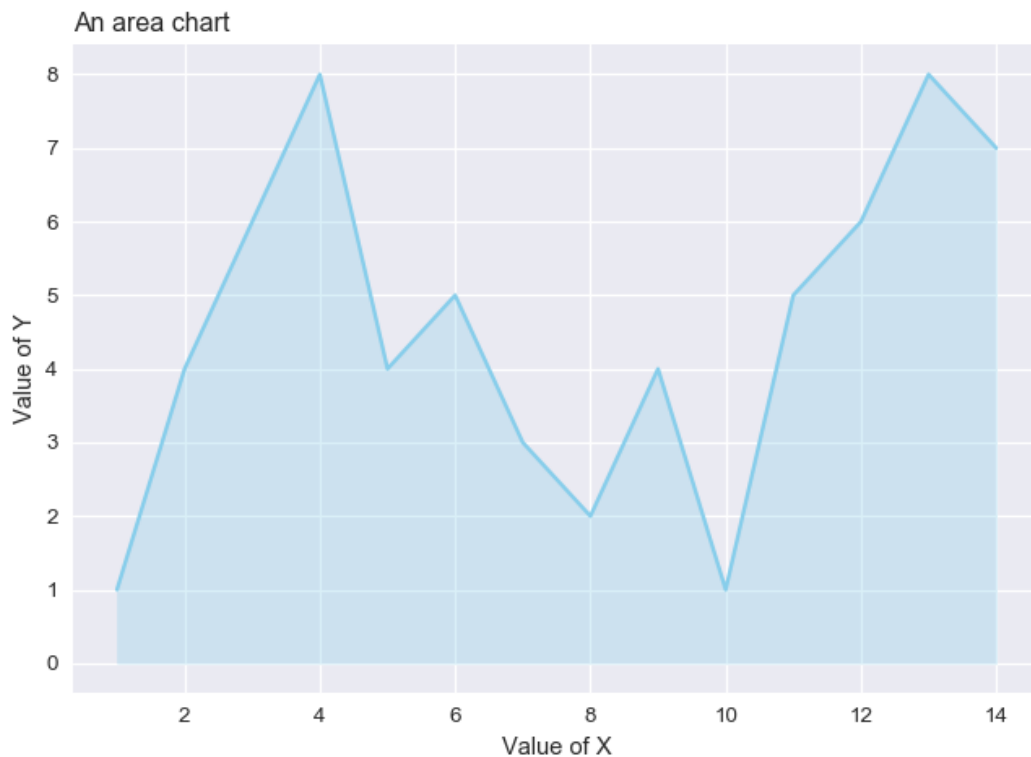
An area chart

g) Stacked Area

- Stacked area charts are best used to show changes in information over time.
- A good example would be the changes in market share among top players or revenue shares by product line over a period of time.
- Stacked area charts might be colourful and fun, but you should use them with attention because they can quickly become a mess.
- Don't stack together more than three to five categories.
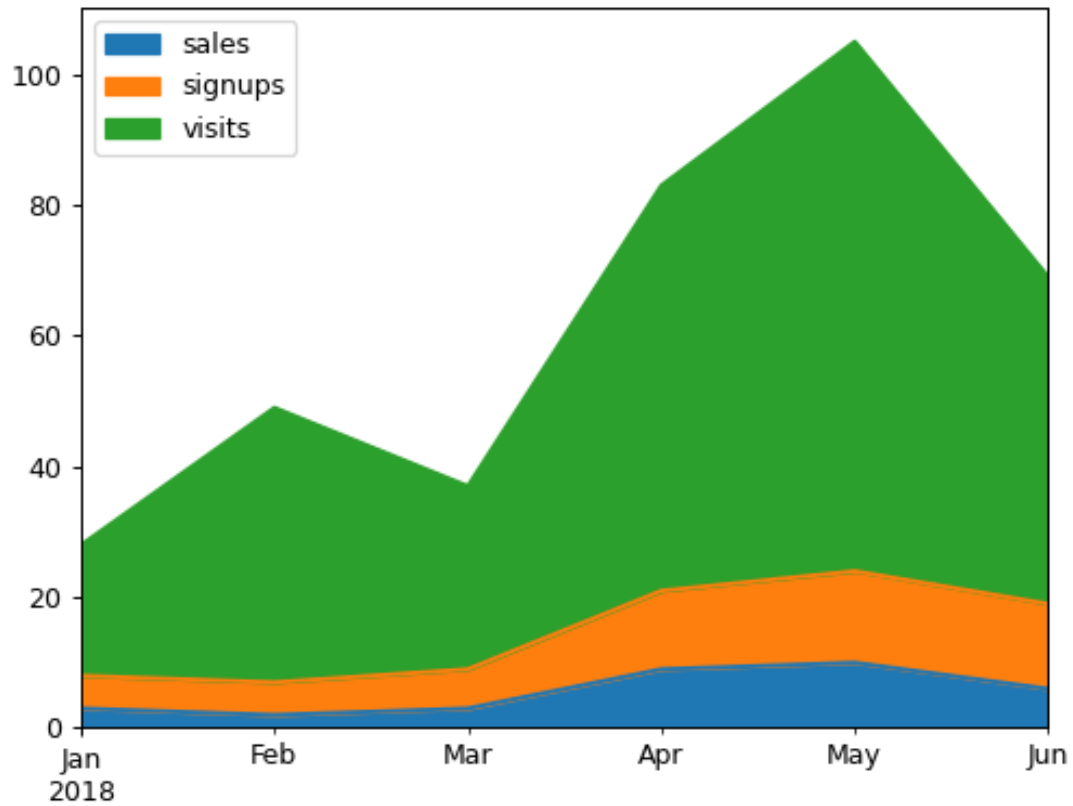
h) Scatter charts

- **Scatter charts are essentially used for correlation and distribution analysis.**
- **Good for showing the relationship between two separate variables where one correlates to another (or doesn't).**
- **Scatter charts can also show the data distribution or clustering trends and help you identify anomalies or outliers.**
- **An example of scatter graphs would be a chart showing total bill vs. tip.**
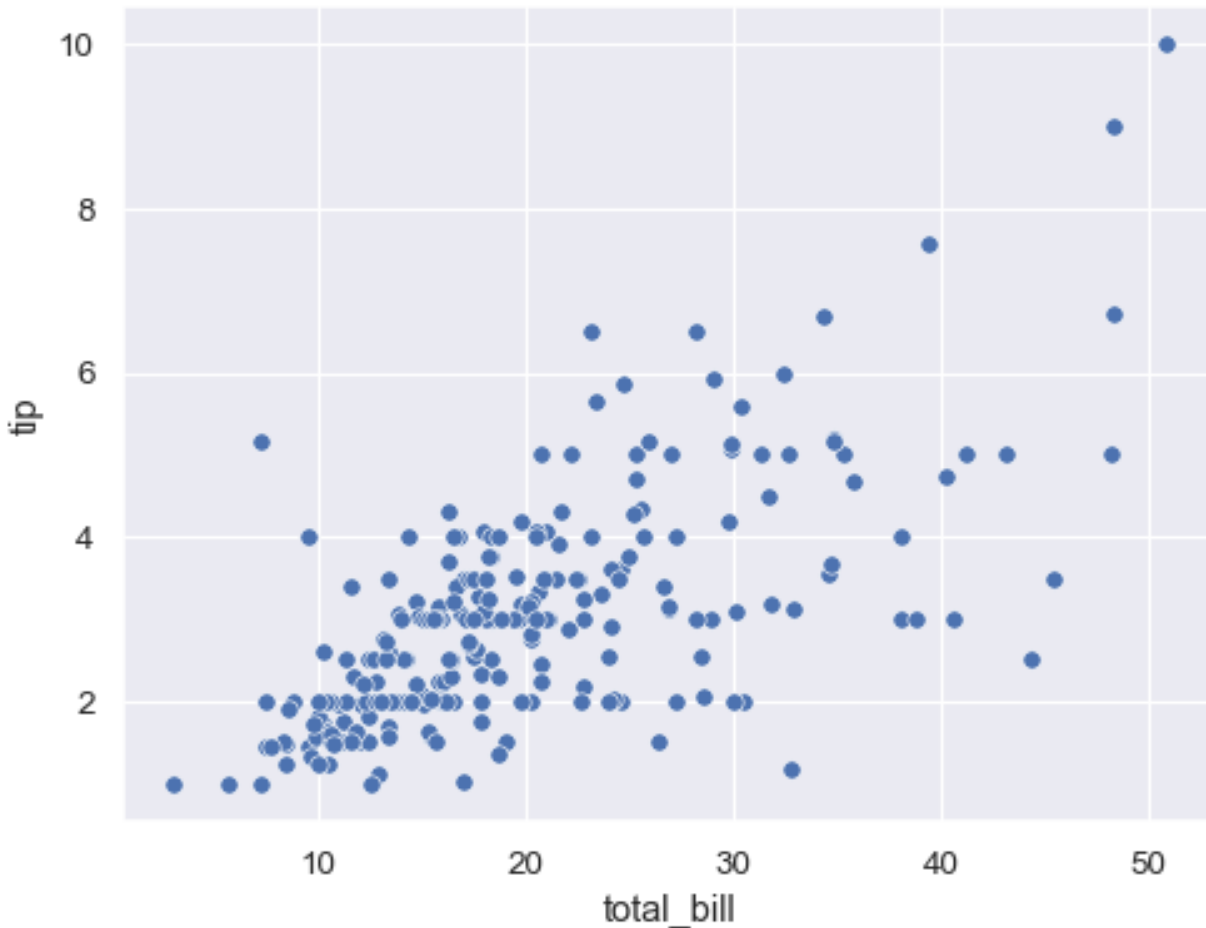
2) Data Visualization do's and don'ts - A general diagnosis

- **Time axis**. When applying time in charts, set it on the horizontal axis. Time should run from left to right. Do not pass over values (time periods), even if there are no values.
- **Proportional values**. The numbers in a chart (displayed as bar, area, or other physically measured elements in the chart) should be directly proportional to the numerical quantities presented.
- **Data-Ink Ratio**. Remove any excess details, lines, colours, and text from a chart that does not add value.
- **Sorting**. For column and bar charts, to enable easier differentiation/comparison, sort your data in ascending or descending order by the value, not alphabetically. This applies also to pie charts.
- You don't need a legend if you have got only one data category.
- **Labels**. Use labels directly on the line, column, bar, pie, etc., whenever possible, to keep away from indirect look-up.
- **Colours**. In any chart, don't use more than six-seven colours.
- **Colours**. For differentiating the same value at different time periods, use the same colour in a different intensity (from light to dark).

- **Colours**. For distinct categories, use different colours. The most widely used colours are white, black, red, green, blue, and yellow.
- **Colours**. Retain the same colour palette or style for all charts in the series, and the same axes and labels for similar charts to make your charts consistent and easy to compare.
- Examine how your charts would look when printed out in grey-scale. If you cannot identify colour differences, you should change the hue and saturation of colours.
- Seven to ten percent of male have a colour deficiency. Keep that in mind when creating charts, making sure they are readable for colour-blind people.
- **Data Complexity**. Don't add too much detail to a single chart. If required, split data into two charts, use highlighting, simplify colours, or change the chart type.

3) Short summary of data visualizations types:

Number chart	It gives a prompt overview of a specific value.
Line Chart	It shows trends & change in data over a period of time.
Waterfall Chart	It demonstrates the static composition of data
Bar Graphs	It is used to compare data of many items
Pie Chart	It indicates the proportional composition of a variable.
Scatter Plot	It is applied to express relations and distribution of large sets of data.
Tables	It shows a large number of precise dimensions and measures.
Area Chart	It portrays a part-to-whole relationship over time.
Bubble Plots	It visualizes 2 or more variables with multiple dimensions.

4) SHORT SUMMARY FOR CHOOSING APPROPRIATE CHART FOR YOUR PROBLEM USING SEABORN

Quantitative / Numerical Variable

Univariate Analysis – Analysing one variable

- displot() – Visualize the distribution of variable (also called histogram)
- boxplot() or violinplot() - To check specifically for outliers

Bivariate Analysis – Relationship between 2 variables

- jointplot() or pairplot()
- lmplot() – Scatter plot with a best fit line

Multivariate Analysis – Relationship between more than 2 variables

- corr() – Correlation matrix followed by
- heatmap() – Visualize the correlation matrix
- pairplot() – Combination of Scatter plots and individual histogram plots for all numerical

variables in the dataset. Also, can assign a categorical variable using hue as an add on)

Qualitative/ Categorical Variables

Univariate Analysis – Analysing one variable

- countplot() – Visualize the distributions of categorical variable

Quantitative vs Qualitative Variables

Analyse how a quantitative variable varies across categorical variable(s)

- boxplot() or violinplot() – To check specifically for outliers
- stripplot() or swarmplot() – Scatter plot across a categorical variable (also helps in checking for outliers)
- barplot() – Can also create a clustered bar chart (assign a categorical variable to hue) or a stacked bar chart (2 bar plots with different colors)
- pointplot()
- lineplot() – Best when looking at trends (Time-related variable along the x-axis)
- catplot() or factorplot() – Analysing a quantitative variable across 2 categorical variables with
- one variable having a high number of categories

5) Important link to explore:

- Uni-Bi-Multivariate analysis.

- [Chart suggestion as a thought-starter](#)

- [Choose the Right Chart Type for Your Data](#)