|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group** | **Plot Type** | **Description** | **Use Case** | **Code Example** |  |
| Distro | Histogram | Shows the distribution of a single numerical variable by dividing it into bins and counting the number of observations in each bin. | Use when you want to see the distribution or frequency of a numerical dataset.  **For comparing distributions of several groups** | sns.histplot(data=data, x='variable', hue='category') |  |
| Distro | KDE Plot | Estimates the probability density function of a continuous random variable. | Useful for visualizing the distribution of a variable, especially when comparing the shapes of distributions. | sns.kdeplot(data=data, x='variable', hue='category', fill=True) |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Categorical | Boxplot | Displays the distribution of a variable based on five summary statistics while highlighting outliers. | Ideal for comparing distributions between several groups or variables.  CONS: unable to show multimodality and clusters | sns.boxplot(x='category', y='value', data=data, hue='subcategory') |  |
| Categorical | Countplot | Shows the counts of observations in each categorical bin using bars. | Best when you need to display the frequency counts of categories in a categorical variable. | sns.countplot(x='category', data=data, hue='subcategory') |  |
| Trend | Lineplot | Plots the data points and connects them with lines, showing the trend of a variable over time or another continuous dimension. | Use to visualize data trends over time or a continuous variable. | sns.lineplot(x='time', y='value', data=data, hue='category') |  |
| Relationship | Scatterplot | Uses dots to represent values from two different variables, one on each axis, showing their relationship. | Effective for examining the relationship or correlation between two numerical variables. | sns.scatterplot(x='value1', y='value2', data=data, hue='category') |  |
| Relationship | Lmplot | Combines scatter plots with regression lines to see the linear relationships between two variables. | Useful for data exploration to understand the relationship between variables and to see if a linear model could fit the data. | sns.lmplot(x='value1', y='value2', data=data, hue='category') |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Relationship | Jointplot | Allows to study the relationship between 2 numeric variables. The central chart displays their correlation. It is usually a scatterplot, a hexbin plot, a 2D histogram or a 2D density plot. | Perfect for a detailed analysis of the relationship between two numerical variables, including their individual distributions along the axes. | Not directly applicable for 'hue', as `jointplot` doesn't support it directly. | kind=”hex”    kind=”kde”    kind=”reg” |
| Categorical | Violinplot | Combines aspects of boxplots with kernel density estimation to show the distribution shape of the data. | Use to compare the distribution of a variable across different categories, especially when the data distribution is multimodal. | sns.violinplot(x='category', y='value', data=data, hue='subcategory') |  |
| Categorical | Stripplot | Represents the distribution of data in a scatterplot format but with a single categorical axis.  Is a scatter plot that differentiates different categoreis | Best for showing the distribution of a numerical variable in small datasets or alongside box/violin plots to show individual observations. | sns.stripplot(x='category', y='value', data=data, hue='subcategory', dodge=True) |  |
| Categorical | Swarmplot | Similar to stripplot, but points are adjusted so they donâ€™t overlap, giving a better representation of the distribution.  Similar to strip plot but avoids overlapping points | Ideal for small to medium-sized datasets where you want to display the distribution of data across categories without overlapping points. | sns.swarmplot(x='category', y='value', data=data, hue='subcategory', dodge=True) |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Composite | Catplot | A figure-level interface for drawing categorical plots onto a FacetGrid. | Versatile for drawing multiple instances of the same plot on different subsets of your dataset, allowing for easy comparison between categories. | sns.catplot(x='category', y='value', data=data, hue='subcategory', kind='bar') | kind=”point” (aka “pinpoint”)    kind=”bar”    kind=”box”    kind=”swarm” |
| Correlation | Heatmap |  | representation of data as a color-encoded matrix.  It is a great way of representing the correlation for each pair of columns in the data. |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Composite | Pairplot | Plots pairwise relationships in a dataset. Creates a grid of Axes such that each variable in data will be shared in the y-axis across a single row and in the x-axis across a single column. | Excellent for exploring correlations between all pairs of numerical variables in a DataFrame. | sns.pairplot(data=data, hue='category') |  |

**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