📊 Graph Concepts Summary

This document serves as a quick reference guide for various types of plots used in data visualization, primarily with the Seaborn library in Python. Each plot is described with its purpose, best use cases, data types, components, and real-world examples.

🔹 Plot Types and Their Uses

1. **Scatter Plot**
    - Shows correlation between two continuous variables.
    - Example: Study hours vs Exam scores.
2. **Box Plot**
    - Displays data distribution through quartiles and highlights outliers.
    - Example: Salary comparisons across departments.
3. **Violin Plot**
    - Similar to a box plot, but includes a KDE for visualizing distribution shape.
    - Example: Test score distributions by gender.
4. **Swarm Plot**
    - Categorical scatter plot that avoids overlap to show all data points.
    - Example: Individual student scores per class.
5. **Histogram**
    - Shows the frequency distribution of a continuous variable.
    - Example: Age distribution of employees.
6. **Bar Plot**

- Compares aggregated statistics (like mean or sum) across categories.
- Example: Revenue per quarter.

7. **Heatmap**
    - Color-coded matrix for visualizing magnitude or correlation values.
    - Example: Correlation matrix for variables.
8. **Density Plot (KDE Plot)**
    - Smooth version of a histogram to estimate distribution.
    - Example: Temperature or GPA distribution.
9. **Joint Distribution Plot**
    - Combines scatter plot with marginal histograms or KDEs.
    - Example: Speed vs Fuel Efficiency.
10. **Factor Plot / Catplot**
    - Multi-category visualization that supports bar, box, etc., via facets.
    - Example: Walking speed by age and gender.

📌 Quick Use-Case Table

Plot Type	When to Use	Data Type
Scatter	Check correlation	Continuous × Continuous
Box	Compare distributions	Continuous × Categorical
Violin	Distribution + Density	Continuous × Categorical
Swarm	Show all points in groups	Continuous × Categorical

Bar	Compare group stats (aggregates)	Aggregated Continuous × Cat.
Histogram	Frequency distribution	Continuous
KDE / Density	Smoothed distribution	Continuous
Heatmap	Matrix visualization	Continuous/Categorical
Jointplot	Joint + Marginal distribution	Continuous × Continuous
Catplot	Multi-faceted group plots	Mixed

📊 Going Beyond Basics: What Else to Know About Graphs ?

1. 🔍 Choosing the Right Visualization

Not all visualizations suit every data type or question. Ask yourself:

- **What is the goal?** Comparison, trend, distribution, relationship?
- **What type of data?** Continuous vs Categorical?
- **How many variables?** Single-variable, bivariate, multivariate?

📘 Advanced Pairings:

Goal	Best Graph Types
Distribution	Histogram, KDE, Violin, ECDF
Comparison (across groups)	Bar, Box, Violin, Strip
Correlation	Scatter, Heatmap, Pairplot, Jointplot

Composition	Stacked Bar, Area Chart, Pie (use sparingly)
Change over time	Line Chart, Area Chart, FacetGrid

2. 🧮 Customization & Clarity

Raw Seaborn/Matplotlib output is rarely presentation-ready. Learn to:

- **Label everything clearly**: titles, axis labels, legends.
- **Use color strategically**: Not just for beauty — to encode meaning.
- **Avoid clutter**: Don't use 3D, pie charts, or too many categories when not needed.
- **Annotation**: Highlight important points or trends directly on the graph.

✅ **Example:**

```
sns.lineplot(data=df, x='year', y='sales', hue='region')
plt.title("Annual Sales by Region")
plt.xlabel("Year")
plt.ylabel("Sales (in millions)")
plt.axvline(2020, linestyle='--', color='gray', label='Pandemic Start')
plt.legend()
```

3. 📐 Multivariate Visualization

When dealing with more than 2 variables:

- **Hue/Color**: Represent a third variable via color.
- **Size**: Use dot size (e.g., in `sns.scatterplot`) to encode a numeric variable.
- **Style**: Line types or markers for categories.
- **Faceting**: Use `sns.FacetGrid` or `sns.catplot(..., col=..., row=...)` to split plots by category.

🎯 **Example:**
sns.relplot(x='age', y='salary', hue='gender', size='experience', data=df, kind='scatter')

4. 📊 Interactive Visualizations

Use tools like **Plotly**, **Altair**, or **Dash** for interactive, explorable plots:

- Tooltips (hover info)
- Zoom/pan
- Dynamic filtering
- Good for dashboards and presentations

Tools:

- `plotly.express`
- `altair`
- `bokeh`
- `dash` (for full dashboards)

5. 🧪 Diagnosing Models with Plots

Graphs are **essential** in machine learning for model evaluation:

Plot Type	Purpose
Residual Plot	Check regression error patterns
ROC/AUC Curve	Classifier performance
Confusion Matrix Heatmap	Classification accuracy
Learning Curve	Model over/underfitting

Feature Importance Bar	Key features in decision models

6. 🧠 Perceptual Principles (Data Viz Best Practices)

- **Avoid pie charts** – human perception of angle/area is poor.
- **Start axes at 0** (especially bar plots) to avoid misleading scales.
- **Consistent color mapping** across related graphs.
- **Don't overload with colors** – especially for categorical data.
- **Use small multiples/facets** instead of crowding multiple lines/plots in one.

7. 📚 Bonus Graph Types to Learn

- **Ridgeline Plot** – stacked KDEs for comparison across groups.
- **ECDF Plot** – better alternative to histograms in some cases.
- **Pairplot** – scatter matrix to explore pairwise relations.
- **Treemap & Sunburst** – for hierarchical data.
- **Network Graphs** – for graph theory, relationships (via NetworkX).