

Seaborn - Notes

Introduction to Seaborn

Seaborn is a Python data visualization library built on top of matplotlib, offering a high-level interface to create visually appealing and statistically informative graphics.

Installation and Importing Libraries

Installation

To install Seaborn, use the following pip command:

```
pip install seaborn
```

Importing Required Libraries

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
```

Using Seaborn Built-in Datasets

Seaborn comes with various built-in datasets useful for demonstration and practice:

```
print(sns.get_dataset_names())

df = sns.load_dataset("tips")

print(df.head())
```

Types of Charts and Their Purposes

1. Bar Plot

Displays aggregated numerical data by categories.

```
sns.barplot(x="day", y="total_bill", data=df)
plt.title("Average Total Bill per Day")
plt.show()
```

2. Count Plot

Shows the count of occurrences for each category.

```
sns.countplot(x="day", data=df)
plt.title("Count of Observations per Day")
plt.show()
```

3. Box Plot

Depicts distribution, central tendency, and variability.

```
sns.boxplot(x="day", y="total_bill", data=df)
plt.title("Boxplot of Total Bill by Day")
plt.show()
```

4. Violin Plot

Combines box plot with kernel density estimation (KDE).

```
sns.violinplot(x="day", y="total_bill", data=df)
plt.title("Violin Plot of Total Bill by Day")
plt.show()
```

5. Strip Plot

Scatter plot for categorical data with jitter to reduce overlap.

```
sns.stripplot(x="day", y="total_bill", data=df, jitter=True)
plt.title("Strip Plot of Total Bill by Day")
plt.show()
```

6. Swarm Plot

Categorical scatter plot without overlapping points.

```
sns.swarmplot(x="day", y="total_bill", data=df)

plt.title("Swarm Plot of Total Bill")

plt.show()
```

7. Histogram

Visualizes the distribution of numerical data.

```
sns.histplot(df["total_bill"], kde=True)

plt.title("Histogram of Total Bill")

plt.show()
```

8. Pair Plot

Plots pairwise relationships between numerical variables.

```
sns.pairplot(df)

plt.suptitle("Pairwise Relationships", y=1.02)

plt.show()
```

9. Heatmap

Displays matrix data, such as correlations, through colors.

```
corr = df.corr()

sns.heatmap(corr, annot=True, cmap="coolwarm")

plt.title("Heatmap of Correlations")

plt.show()
```

Fundamental Parameters

- data: Dataset for plotting (DataFrame).
- x, y, hue: Variables for axes and color grouping.
- palette: Defines color schemes.
- order: Category order.
- ci: Confidence interval (e.g., ci='sd').
- kind: Specifies plot type in advanced functions like relplot, catplot.

This detailed guide provides foundational knowledge on Seaborn for data visualization in Python.