Seaborn Class Theory Notes (Full Detailed)

Interquartile Range (IQR) and Outliers:

The Interquartile Range (IQR) measures statistical dispersion and is calculated by subtracting Q1 from Q3:

- Q1 (25th percentile): 25% of data points fall below Q1.
- Q2 (Median, 50th percentile): Middle of dataset.
- Q3 (75th percentile): 75% of data points fall below Q3.
- IQR Formula: IQR = Q3 - Q1

Lower and Upper Fence for Outliers:
- Lower Fence = Q1 - 1.5 * IQR
- Upper Fence = Q3 + 1.5 * IQR
- Outliers: Data points below Lower Fence or above Upper Fence.

Correlation vs Covariance:

Correlation:
- Measures strength and direction of a linear relationship between two variables.
- Ranges between -1 and +1.
- +1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, 0 indicates no correlation.
- Unit-less measure.

Covariance:
- Measures how two variables vary together.
- Positive covariance: Variables move together.
- Negative covariance: Variables move inversely.
- No fixed range, depends on units of variables.
- Units dependent.

Difference:
- Correlation is standardized (unit-free) and bounded between -1 and +1.
- Covariance is not standardized and depends on data units.

Distribution Types:

---

1. Normal Distribution:

- Symmetrical, bell-shaped.
- Mean = Median = Mode.

2. Positively Skewed Distribution (Right Skewed):

- Tail extends to the right.
- Mean > Median.

3. Negatively Skewed Distribution (Left Skewed):

- Tail extends to the left.
- Mean < Median.

4. Uniform Distribution:

- All outcomes equally likely.
- Flat distribution.

Seaborn Styling & Customizations:

- Themes: sns.set_style('whitegrid', 'darkgrid', 'white', 'dark', 'ticks')
- Context for plot scaling: sns.set_context('notebook', 'talk', 'poster', 'paper')
- Palettes: sns.color_palette('deep', 'muted', 'bright', 'pastel', 'dark', 'colorblind', 'magma', 'coolwarm')
- Customizing Plots:

```
plt.title('Title')
plt.xlabel('X-axis label')
plt.ylabel('Y-axis label')
plt.xticks(rotation=90)
```

- Saving plots:

```
plt.savefig('plot_name.png')
```

Why Data Visualization?

- Easily identify patterns, trends, relationships, and outliers.
- Better communication of insights from data.
- Supports decision making by visual data interpretation.

Choosing the Right Plot:

- Distribution: Histogram, KDE, Boxplot, Violinplot.
- Relationship between two variables: Scatterplot, Regression plot.
- Categorical data: Countplot, Barplot.
- Multi-variable relationships: Pairplot, Heatmap.
- Time-series/Trend analysis: Lineplot, FacetGrid.