Lecture Notes:
Matplotlib and Seaborn

Key Definitions & Concepts

- **Matplotlib (plt)**: A low-level Python library that gives fine control over every plot element
- **Seaborn (sns)**: A high-level wrapper over Matplotlib designed for **statistical visualisation**
- **Figure**: The overall plot canvas
- **Axes**: Subplots or coordinate spaces within the figure
- **Tick Labels**: Numbers or categories on x- and y-axes
- **Hues**: Used in Seaborn to distinguish data by categories using colour
- **Matplotlib** gives precise control over plots, while **Seaborn** provides pre-built, declarative interfaces for clean visualisations
- Plot customisation, like `xticks`, `ylabel`, and `legend,` help in **communicating insights** clearly
- `hue` in Seaborn maps categories to colours, allowing **groupwise comparison**
- **Boxplots** summarise distributions using **median, IQR, and whiskers**; **histograms** show distributions using **bins**

upGrad

Implementation

Plot Definitions in Matplotlib vs Seaborn

Task	Matplotlib (`plt`)	Seaborn (`sns`)

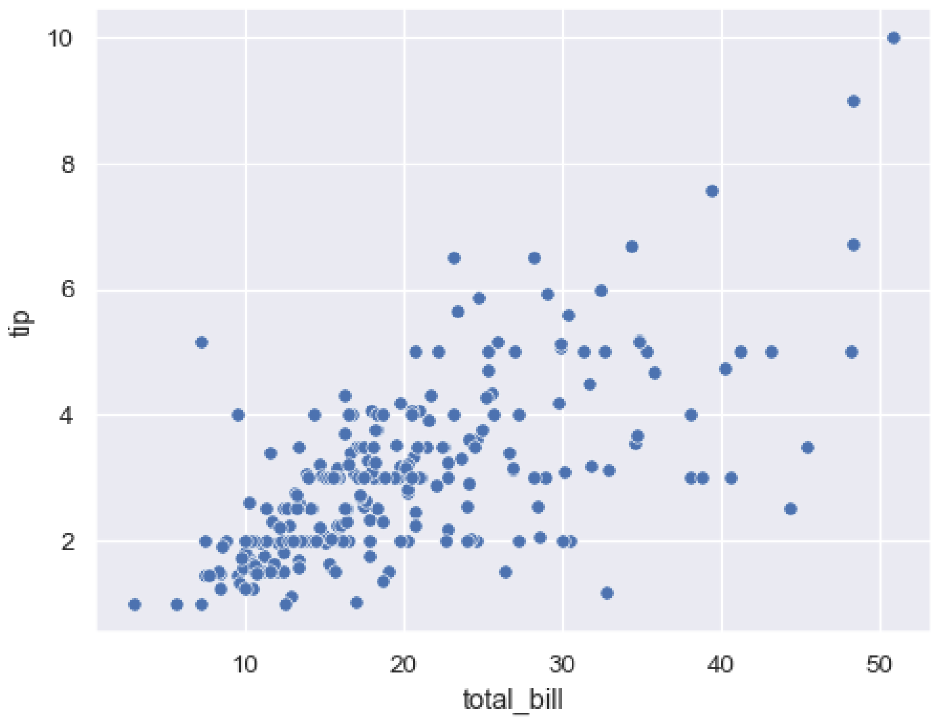
Create Plot	`plt.plot(x, y)`
	Basic line plot, minimal defaults; requires manual formatting
		`sns.lineplot(x=x, y=y, data=df)`
		Styled by default;
		Works seamlessly with DataFrames

Set Title	`plt.title("Title")` Sets the title on the current figure	`ax.set_title("Title")` Used with `sns` plots inside subplots; or `plt.title()`
Axis Labels	`plt.xlabel("X")`, `plt.ylabel("Y")` Full manual control	Same as `plt`; Seaborn does not override axis label behaviour
Legend	`plt.legend()` Manual labels required (via `label=` in plot calls)	Automatically generates from `hue` or `label=`; Use `plt.legend()` to customise
Tick Labels	`plt.xticks()`, `plt.yticks()` Precise tick control (values, labels, rotation)	Inherits from Matplotlib; commonly used with `plt.xticks(rotation=45)`
Save Plot	`plt.savefig("figure.png")` Saves the current figure as an image	Same; Seaborn uses Matplotlib's backend for rendering and saving
Hue (Colour Grouping)	Manual Use `color=`, `label=` for each group or series manually	Automatic Use `hue='column'` to group and colour data by a categorical variable
Styling & Themes	Requires manual setting: colours, grid, font, etc. via `plt.style.use()`	Comes with built-in themes (`sns.set_theme()`); Consistent aesthetics out-of-the-box

upGrad

Plot Types

Scatter Plot



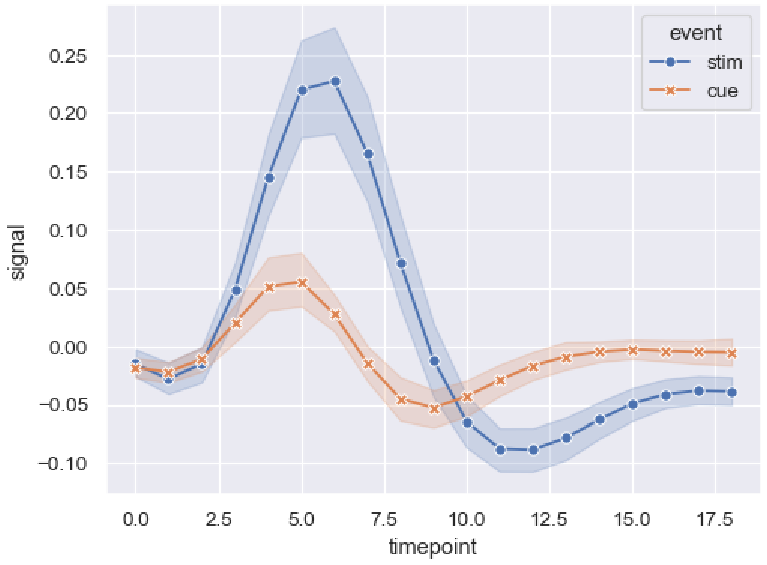
Matplotlib

```python
plt.scatter(x='Age', y='Salary', data=df, color='blue',
label='Employees')
plt.xlabel('Age')
plt.ylabel('Salary')
plt.legend()
```

Seaborn

```python
sns.scatterplot(x='Age', y='Salary', data=df, hue='Department')
```

Line Plot



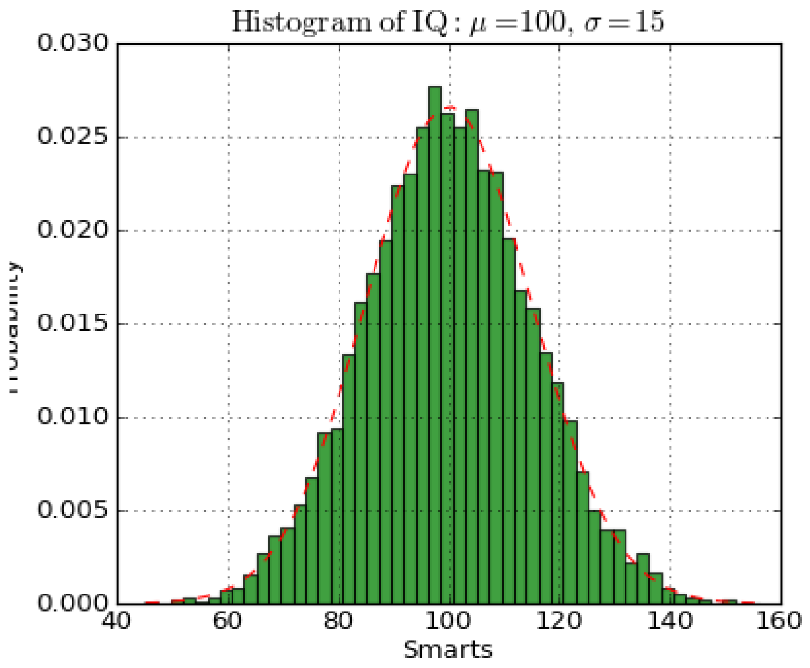
Matplotlib

```
Python

plt.plot(df['Month'], df['Sales'], label='Sales')
plt.xticks(rotation=45)
plt.xlabel('Month')
plt.ylabel('Sales')
plt.legend()
```

Seaborn

```
Python

sns.lineplot(x='Month', y='Sales', data=df, hue='Region')
```

Histogram



Matplotlib

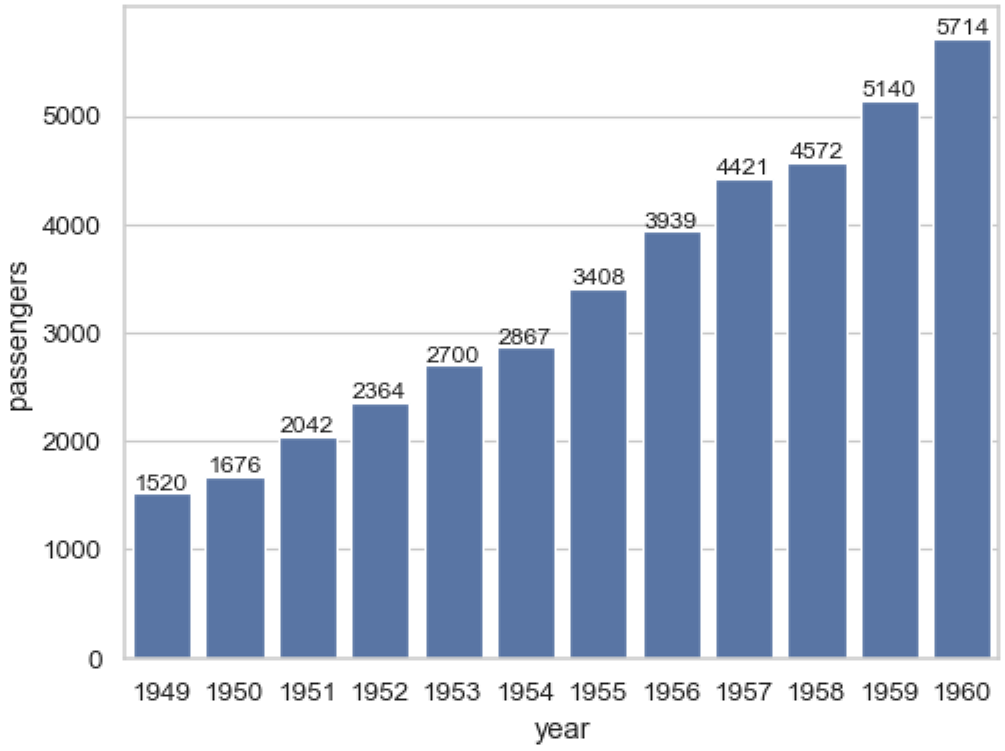
```
Python

plt.hist(df['Marks'], bins=10, color='skyblue',
edgecolor='black')
```

Seaborn

```
Python

sns.histplot(df['Marks'], bins=10, kde=True)
```

`bins` define the granularity of distribution; `kde=True` overlays a smooth curve.

Bar Plot



Matplotlib

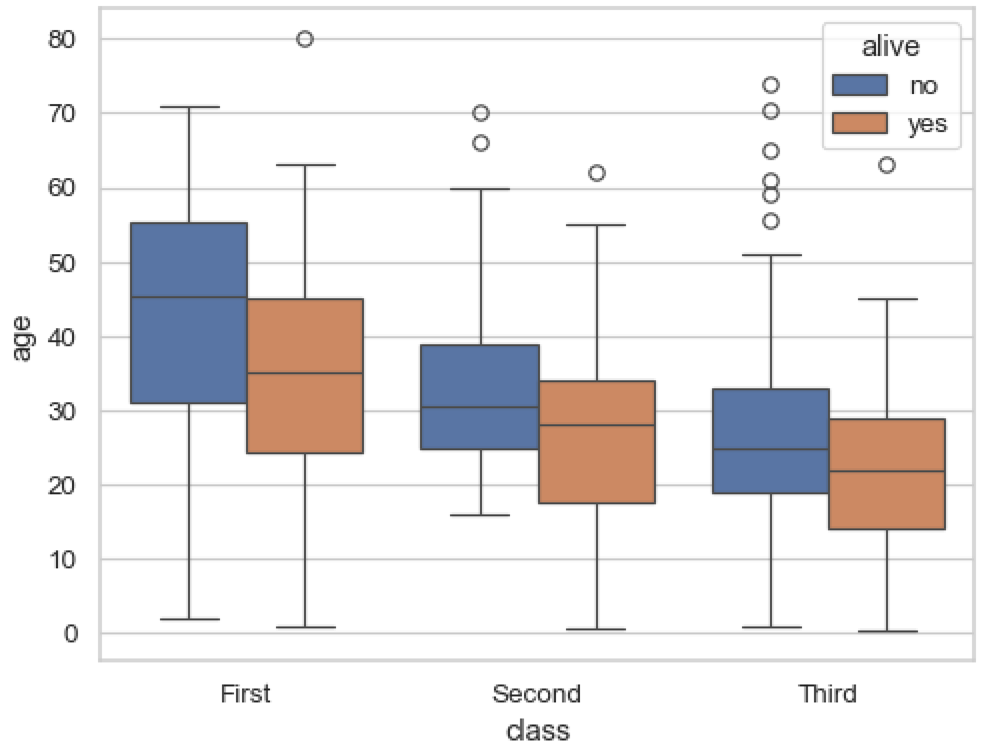
```python
categories = df['Department'].value_counts()

plt.bar(categories.index, categories.values)
```

Seaborn

```python
sns.barplot(x='Department', y='Salary', data=df,
estimator='mean')
```

Seaborn allows aggregation (`sum`, `mean`, `count`) directly.

Box Plot



Seaborn

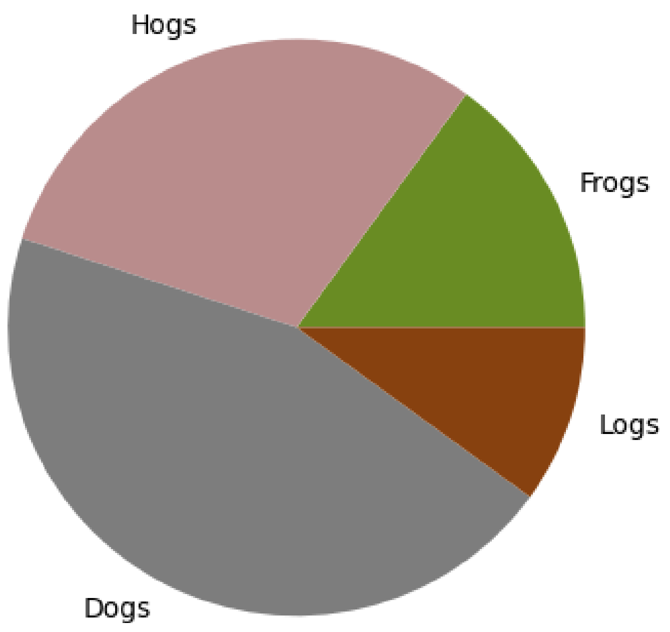
Python

```python
sns.boxplot(x='Department', y='Salary', data=df)
```

Interpretation:

- **Median**: Central line in the box
- **IQR**: Height of the box (Q3 - Q1)
- **Whiskers**: Extend to 1.5×IQR
- **Outliers**: Points outside the whiskers

Pie Chart (Matplotlib only)



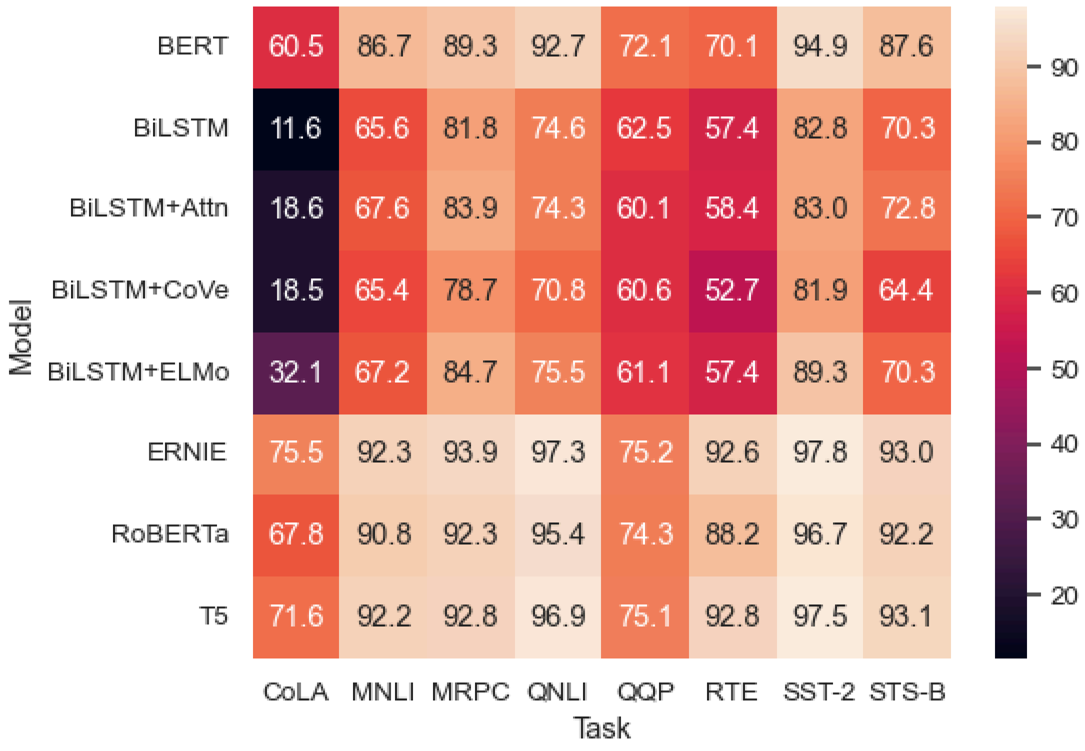
```python
sizes = [40, 30, 20, 10]
labels = ['A', 'B', 'C', 'D']
explode = [0.1, 0, 0, 0]
plt.pie(sizes, labels=labels, explode=explode, autopct='%1.1f%%')
```

We could add annotations to show the percentage of each slice of the pie chart using `autopct` as well as explode a particular slice for emphasis using `explode`.

`explode` takes in a sequence where each element corresponds to a slice in your pie chart

- A value of 0 at an index means that the corresponding slice will not be exploded.
- A value greater than 0 (e.g., 0.1, 0.2) at an index will explode that slice by the specified fraction of the radius. A larger value will result in a greater separation.

Heatmap



	CoLA	MNLI	MRPC	QNLI	QQP	RTE	SST-2	STS-B
BERT	60.5	86.7	89.3	92.7	72.1	70.1	94.9	87.6
BiLSTM	11.6	65.6	81.8	74.6	62.5	57.4	82.8	70.3
BiLSTM+Attn	18.6	67.6	83.9	74.3	60.1	58.4	83.0	72.8
BiLSTM+CoVe	18.5	65.4	78.7	70.8	60.6	52.7	81.9	64.4
BiLSTM+ELMo	32.1	67.2	84.7	75.5	61.1	57.4	89.3	70.3
ERNIE	75.5	92.3	93.9	97.3	75.2	92.6	97.8	93.0
RoBERTa	67.8	90.8	92.3	95.4	74.3	88.2	96.7	92.2
T5	71.6	92.2	92.8	96.9	75.1	92.8	97.5	93.1

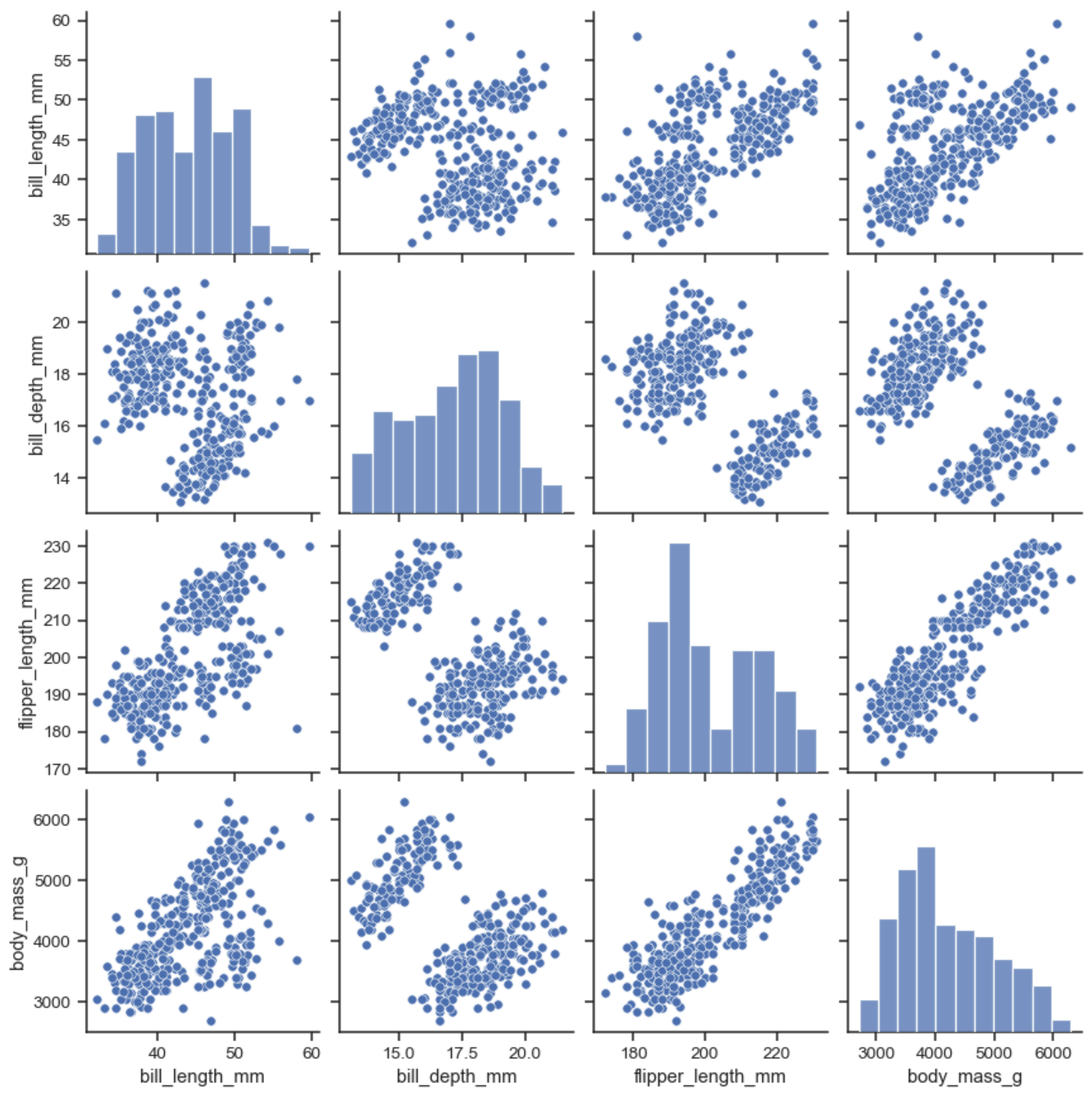
Seaborn

Python

```python
corr = df.corr()

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

Common use: show a **correlation matrix** for numeric variables.

Pairplot



Seaborn

```python
sns.pairplot(df, hue='Species')
```

Shows pairwise scatter plots + histograms for all numerical columns. Excellent for multivariate inspection.

Summary

Plot Type	Use Case	Key Argument(s)
`scatterplot`	Relationship between two variables	`x`, `y`, `hue`, `style`
`lineplot`	Trends over ordered data (e.g., time)	`x`, `y`, `hue`, `ci`
`histplot`	Distribution of a numeric variable	`bins`, `kde`, `hue`
`barplot`	Categorical mean/counts	`estimator`, `ci`, `hue`
`boxplot`	Distribution + outliers	`x`, `y`, `hue`
`pie`	Part-to-whole (not in Seaborn)	`labels`, `autopct`, `explode`
`heatmap`	Matrix/Correlation visualisation	`annot`, `cmap`, `vmin`, `vmax`
`pairplot`	Explore multivariate numeric distributions	`hue`, `palette`, `kind`

Pitfalls

- `hue` only works in **Seaborn** (not Matplotlib)
- Too many categories in pie/bar → cluttered visuals
- Histogram interpretation depends on **bin size**
- Pairplots can be slow on large datasets; sample first
- Misaligned axis labels or missing legends reduce readability

Additional Reading

- Seaborn Documentation
- Matplotlib Documentation
- Seaborn Examples Gallery