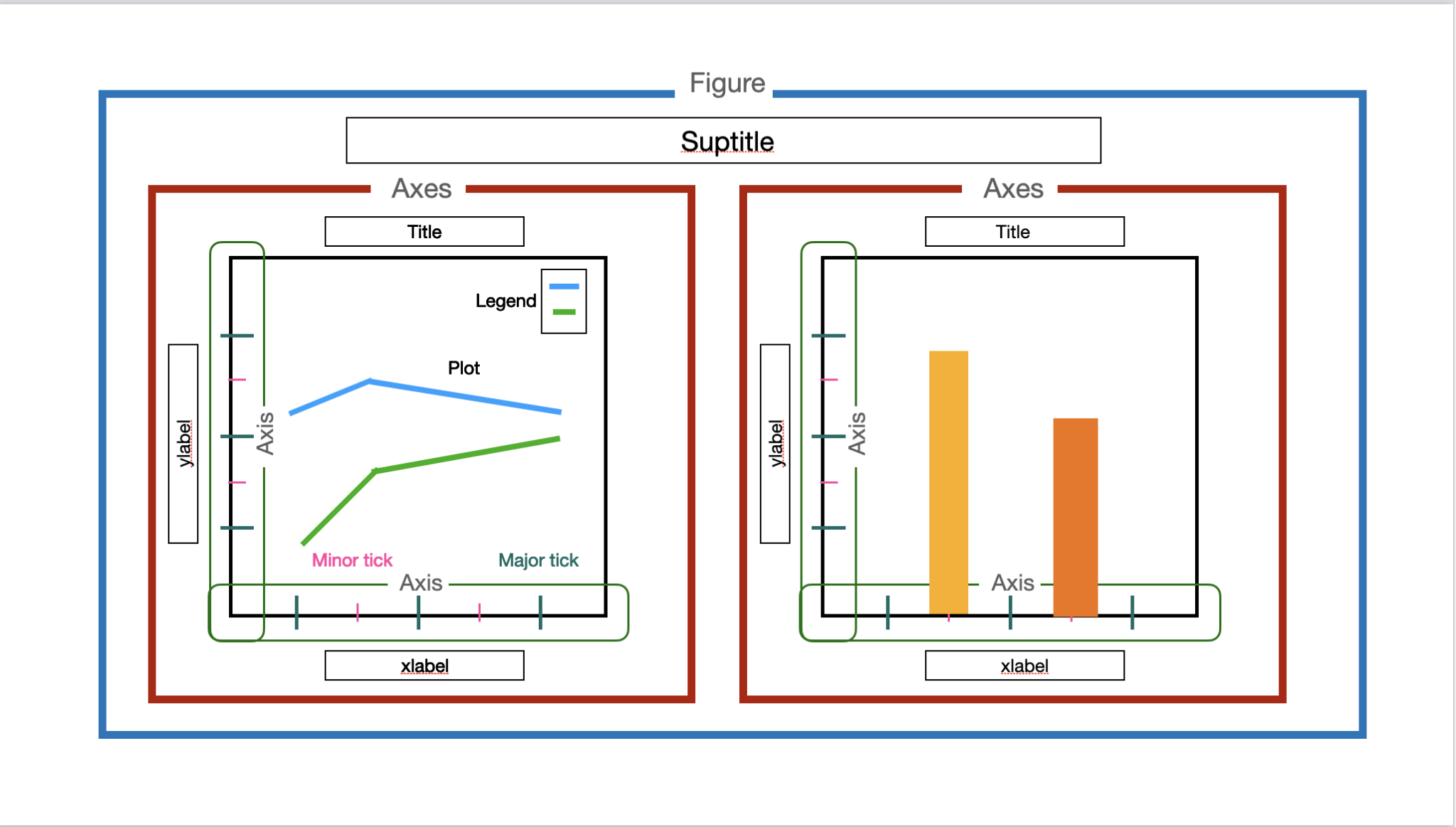
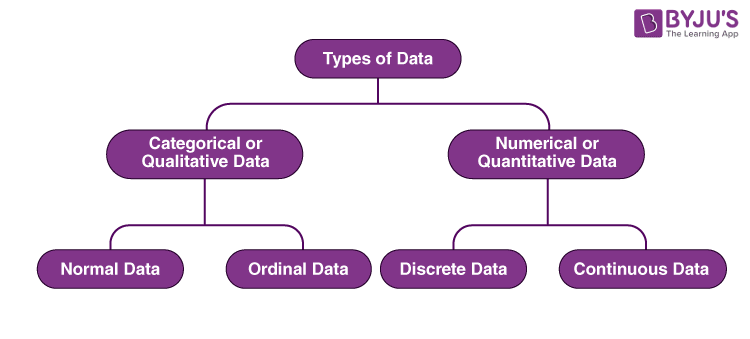
**Content:**

What is a plot?

* 1. **Anatomy**
  2. **Major components:**
* **Figure**: The overall window or page that everything is drawn on.
* **Axis**: Simply the `x-axis` and `y-axis`
* **Axes**: It is the area on which the data is plotted
* **x-label**: Name of x-axis
* **y-labe**l: Name of y-axis
* **Major ticks**: subdivides the axis into major units.
* **Minor ticks**: subdivides major tick units.
* **Title**: Title of each plot (Axes)
* **Legend**: describes the elements in the plot, blue and green curves in this case
* **Suptitle**: The common title of all the plots

| **Figure** | plt.figure(figsize=(x,y))  E.g.  plt.figure(figsize=(15,10)) |
| --- | --- |
| **Labels** | plt.xlabel(‘x label name’)  plt.ylabel(‘y label name’) |
| **Title** | plt.title(‘Title of the plot) |
| **Suptitle** | fig.suptitle(‘Title of the whole figure’)  # Used in case of subplots |
| **Ticks** | plt.xticks(rotation=90)  # For x ticks  plt.yticks(rotation=90)  # For y ticks |
| **Legend** | plt.legend() |

1. **Types of Data:**
   * 1. **Numerical data:**
        1. **Discrete** -> Can only take finite values (E.g. Year, 2010, 2011, etc.)
        2. **Continuous** > Can take any numerical value (E.g. temperature, pressure, etc.)
     2. **Categorical data:**
        1. **Ordinal**: Categorical Data with an order (E.g. low, medium, high)
        2. **Non-ordinal/nominal**: Categorical Data without any order (example gender as Male/Female)



1. **Installing and importing matplotlib and seaborn**

| **Installing** | pip install matplotlib  pip install seaborn |
| --- | --- |
| **Importing convention** | import matplotlib.pyplot as plt  import seaborn as sns |

1. **Univariate Data Analysis:**
   1. **Categorical Data**

| **Countplot** (Type of barplot) | sacountplot(x=’col\_name’, data=df)  E.g.  sns.countplot(x=’Genre’,data=df) |
| --- | --- |
| **Pie chart** | y = np.array([35, 25, 25, 15])  mylabels = ["Toyota", "Suzuki", "Mercedes", "Buggati"]  plt.pie(y, labels = mylabels) |

* 1. **Numerical Data**

| **Histogram** | sns.histplot(x=’col\_name’, data=df)  E.g. sns.histplot(x=’Year’, data=df) |
| --- | --- |
| **KDE Plot** | sns.kdeplot(x=’col\_name’, data=df)  E.g. sns.kdeplot(x=’Year’, data=df) |
| **Boxplot** | sns.boxplot(x=’col\_name’, data=df)  E.g. sns.boxplot(x=’Global\_sales’, data=df) |

1. **Bivariate**
   1. **Numerical-Numerical**

| **Line Plot** | sns.lineplot(x=’xcol’, y=’ycol’, data=df)  E.g.  sns.lineplot(x='Year', y='NA\_Sales', data=data) |
| --- | --- |
| **Scatter Plot** | sns.scatterplot(x=’xcol’, y=’ycol’, data=df)  E.g.  sns.scatterplot(x='Rank', y='Global\_Sales, data=df') |

* 1. **Categorical-Categorical**

| **Dodged barplot**  (Extension of countplot) | sns.countplot(x=’xcol’, hue=’hueCol’, data=df)  E.g.  sns.countplot(x='Publisher',hue='Platform',data=plat\_data) |
| --- | --- |
| **Stacked barplot** | sns.countplot(x=’xcol’, hue=’hueCol’, data=df, dodge=False)  E.g.  sns.countplot(x='Publisher',data=top3\_data, hue='Genre',dodge=False) |

* 1. **Categorical-Numerical**

| **Boxplot** | sns.boxplot(x=’xcol’, y=’ycol’, data=df)  E.g.  sns.boxplot(x='Publisher', y='Global\_Sales', data=pub\_data) |
| --- | --- |
| **Barplot** | sns.barplot(x=’xcol’, y=’ycol’, data=df)  E.g.  sns.barplot(x="Genre", y="Global\_Sales", data=sales\_data) |

1. **Multivariate**

| **Numerical Numerical Categorical**  (Scatterplot with colour) | sns.scatterplot(x=’xcol’, y=’ycol’, hue=’hueCol’, data=df)  E.g.  sns.scatterplot(x='NA\_Sales', y='EU\_Sales',hue='Publisher',data=top3\_data) |
| --- | --- |
| **Categorical Categorical Numerical** (Dodged boxplot) | sns.boxplot(x=’xcol’, y=’ycol’, hue=’hueCol’, data=df)  Eg.  sns.boxplot(x='Publisher',y='Global\_Sales',hue='Genre',data=top3\_data) |
| **Numerical-Numerical-Numerical** (Bubble Chart) | sns.scatterplot(x=’xcol’, y=’ycol’, size=’rankCol’, sizes=tuple\_of\_sizes, data=data)  E.g.  sns.scatterplot(x='NA\_Sales', y='JP\_Sales', size='Rank', sizes=(1, 200), data=data) |

1. **Subplots**

We can plot multiple plots in a single figure

plt.figure()

plt.subplot(row,col,1)

# plot

…

E.g.

plt.figure(figsize=(20,12)).suptitle("NA Sales vs regions",fontsize=20)

# Using a 2x3 subplot

plt.subplot(2, 3, 1)

sns.scatterplot(x='NA\_Sales', y='EU\_Sales', data=top3\_data)

plt.subplot(2, 3, 3)

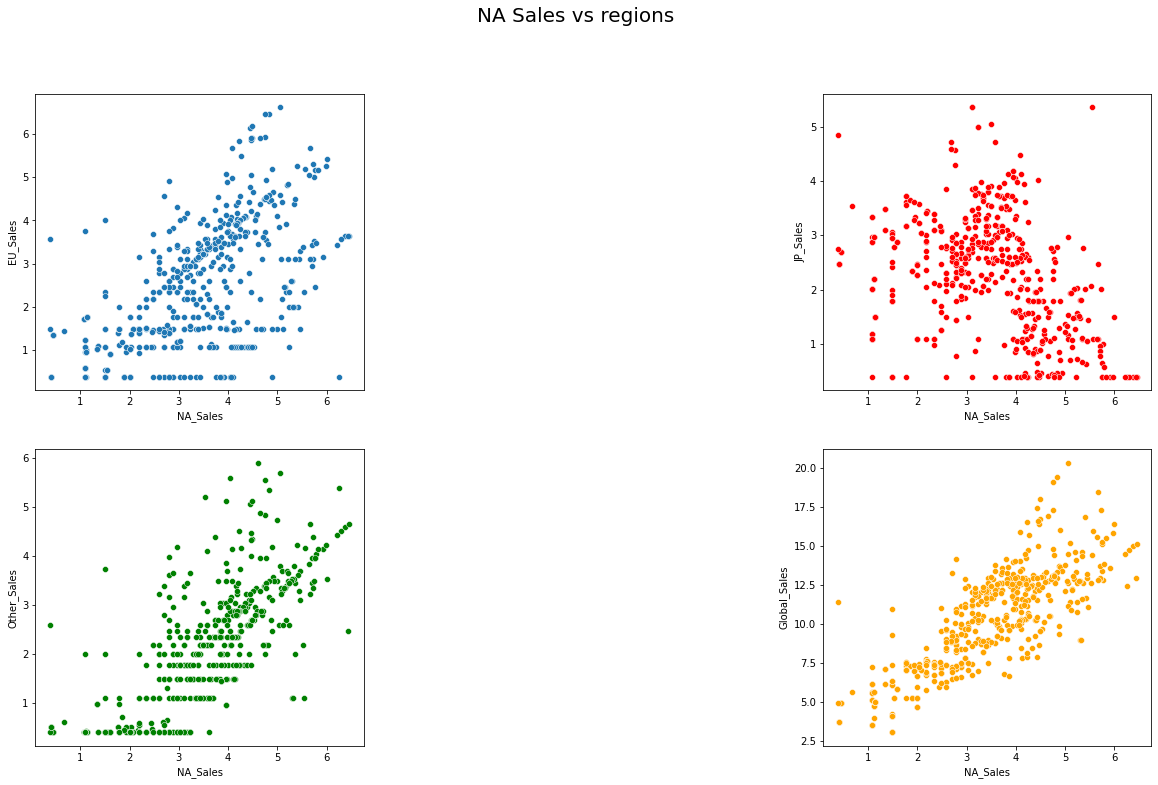
sns.scatterplot(x='NA\_Sales', y='JP\_Sales', data=top3\_data, color='red')

plt.subplot(2, 3, 4)

sns.scatterplot(x='NA\_Sales', y='Other\_Sales', data=top3\_data, color='green')

plt.subplot(2, 3, 6)

sns.scatterplot(x='NA\_Sales', y='Global\_Sales', data=top3\_data, color='orange')



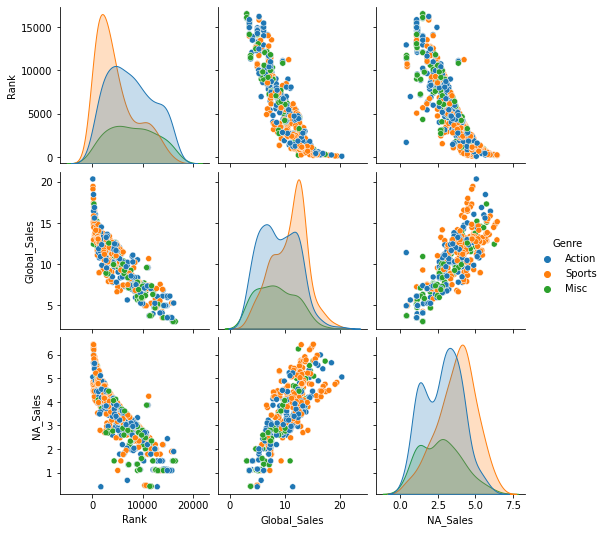
1. **Pairplot**

Displays a scatterplot for each pair of attributes, can provide a hue for each category too

sns.pairplot(data=df, hue=’hueCol’)

E.g.

sns.pairplot(data=df, hue=’Genre’)



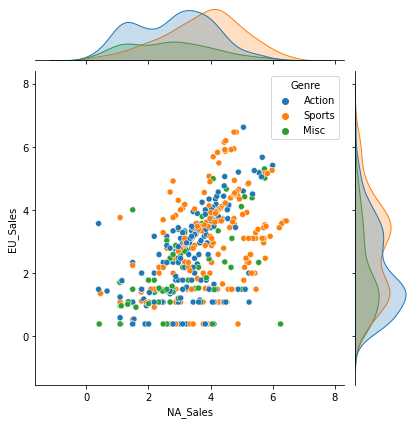
1. **Jointplot**

Draws multiple types of plot of two variables in the same plot

sns.jointplot(x=’xcol’, y=’ycol’, data=data, hue=’hueCol)

E.g.

sns.jointplot(x='NA\_Sales', y='EU\_Sales', data=top3\_data, hue=’Genre’)



1. **Correlation and Heatmaps**

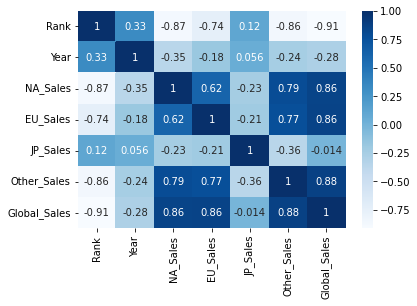
Plot a heatmap of correlation between various variables

sns.heatmap(df.corr(), cmap=’colour\_map)

E.g.

sns.heatmap(top3\_data.corr(), cmap= "Blues", annot=True)

# plots a heatmap of the data with the correlation coefficient values annotated



1. **Plotting with Pandas**

| **Lineplot** | df.plot(x=’col1’, y=’col2’)  E.g.  df.plot(x="Rank", y = "Median") |
| --- | --- |
| **Histogram** | df[‘col’].plot(kind=’hist’)  E.g.  df['Median'].plot(kind="hist") |
| **Barplot** | df[‘col’].value\_counts().plot(kind=’bar’)  E.g.  df['Major\_category'].value\_counts().plot(kind = 'bar') |
| **Scatterplot** | df.plot(x=’col1’, y=’col2’, kind=’scatter’)  E.g.  df.plot(x="Median", y="Unemployment\_rate", kind="scatter") |