Data Visualization using matplotlib

S.M. Kia

Data Visulization using matplotlib

❖ An introduction to matplotlib

❖ Anatomy of a figure

   ❖ What is a figure?

   ❖ What is an axis?

❖ Attributes of a plot in matplotlib

❖ Several axes in the same figure using subplots

❖ Common plot types for data visualization

   ❖ Simple plot, Scatter plot, Pie chart, Bar plot, Histogram, boxplot

❖ Matplotlib for Pandas data frames

Matplotlib

❖ Matplotlib is a Python package for data visualization on NumPy arrays.

❖ It can be also used with Pandas data frames (sometimes it is not that efficient).

❖ pyplot is an API for data visualization tools in matplotlib:

```python
import matplotlib.pyplot as plt
```

Anatomy of a Figure in Matplotlib

❖ A figure is the basic object for plotting data in Python. When we plot data on a graph, that graph will be drawn on a figure.

❖ An axes object is a region within the figure with the data space. A given figure can contain multiple axes, but a given axes object can only be in one figure.

TILBURG ◆ UNIVERSITY

4
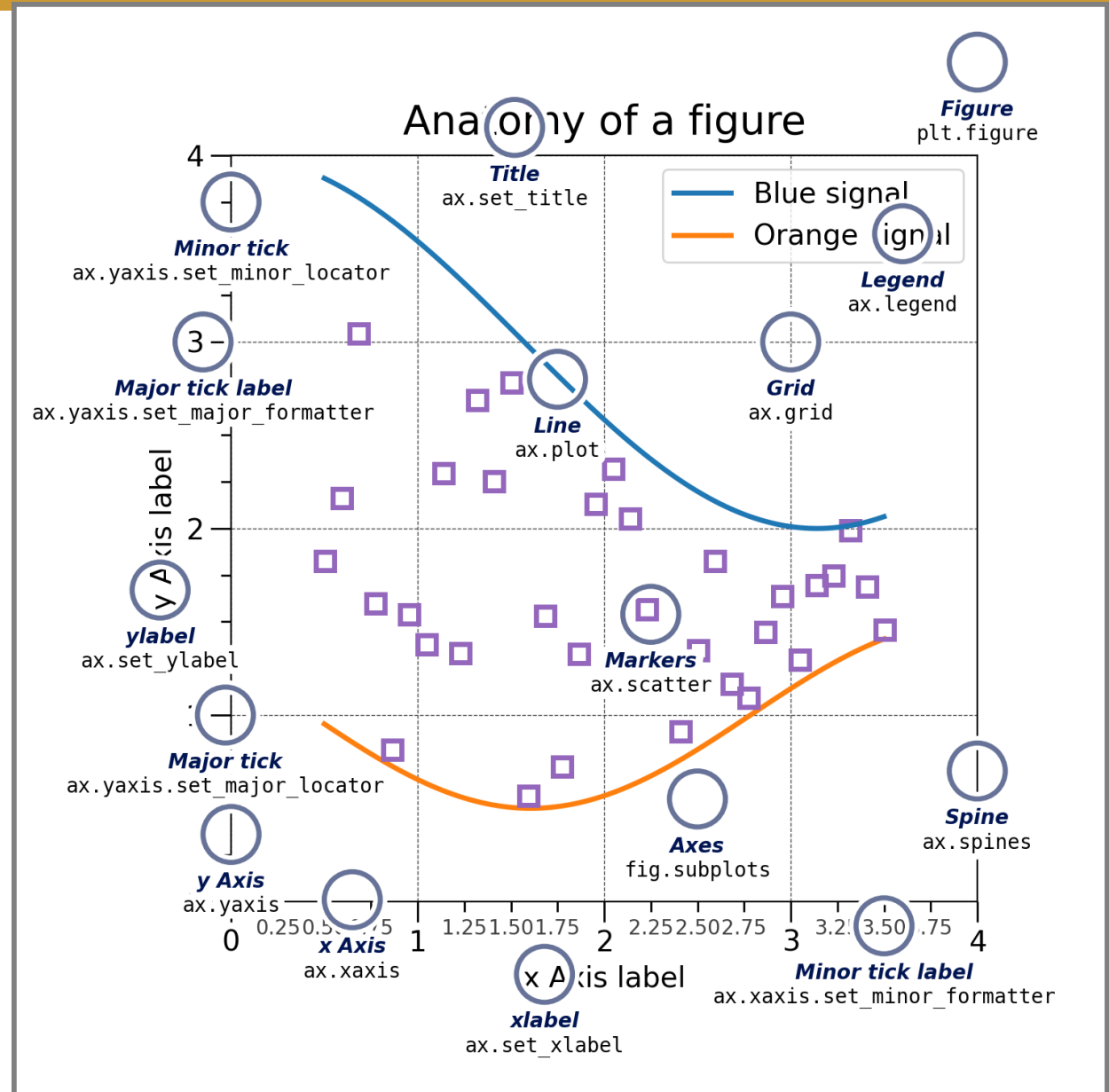
Attributes of a Plot in Matplotlib

❖ Color (c or color)

https://www.rapidtables.com/web/color/RGB_Color.html

character	color
'b'	blue
'g'	green
'r'	red
'c'	cyan
'm'	magenta
'y'	yellow
'k'	black
'w'	white

Color	Color Name	Hex Code #RRGGBB	Decimal Code R,G,B
	maroon	#800000	(128,0,0)
	dark red	#8B0000	(139,0,0)
	brown	#A52A2A	(165,42,42)
	firebrick	#B22222	(178,34,34)
	crimson	#DC143C	(220,20,60)
	red	#FF0000	(255,0,0)
	tomato	#FF6347	(255,99,71)
	coral	#FF7F50	(255,127,80)
	indian red	#CD5C5C	(205,92,92)
	light coral	#F08080	(240,128,128)
	dark salmon	#E9967A	(233,150,122)
	salmon	#FA8072	(250,128,114)
	light salmon	#FFA07A	(255,160,122)
	orange red	#FF4500	(255,69,0)
	dark orange	#FF8C00	(255,140,0)
	orange	#FFA500	(255,165,0)
	gold	#FFD700	(255,215,0)

Attributes of a Plot in Matplotlib

❖ Line style (linestyle): see the list at
https://matplotlib.org/stable/gallery/lines_bars_and_markers/linestyles.html

- A string:

linestyle	description
'-' or 'solid'	solid line
'--' or 'dashed'	dashed line
'-.' or 'dashdot'	dash-dotted line
':' or 'dotted'	dotted line
'none', 'None', ' ', or ''	draw nothing

❖ Line width (linewidth): Sets the line width in points

❖ Opacity (alpha): Opacity scale in 0 and 1.0 range.

❖ Label (label): Label for a specific draw in the legend

TILBURG ◆ UNIVERSITY

# Attributes of a Plot in Matplotlib

❖ Marker (marker, see the list at https://matplotlib.org/stable/api/markers_api.html), marker size (markersize), marker face (markerfacecolor) and edge color (markeredgecolor):

marker	symbol	description
"."	•	point
","	·	pixel
"o"	●	circle
"v"	▼	triangle_down
"^"	▲	triangle_up
"<"	◄	triangle_left
">"	►	triangle_right
"1"	Y	tri_down
"2"	⅄	tri_up
"3"	⊰	tri_left

marker	symbol	description
"4"	⊱	tri_right
"8"	●	octagon
"s"	■	square
"p"	⬠	pentagon
"P"	✚	plus (filled)
"*"	★	star
"h"	⬡	hexagon1
"H"	⬢	hexagon2
"+"	+	plus
"x"	✕	x
"X"	✖	x (filled)

marker	symbol	description
"D"	◆	diamond
"d"	◆	thin_diamond
"|"	|	vline
"_"	—	hline
0 (TICKLEFT)	—	tickleft
1 (TICKRIGHT)	—	tickright
2 (TICKUP)	|	tickup
3 (TICKDOWN)	|	tickdown
4 (CARETLEFT)	◄	caretleft
5 (CARETRIGHT)	►	caretright
6 (CARETUP)	▲	caretup
7 (CARETDOWN)	▼	caretdown

Annotating an Axes

❖ Title (plt.title): a string specifying the title for an axis in a figure

❖ Label of x-axis (plt.xlabel): a string specifying the label for x-axis

❖ Label of y-axis (plt.ylabel): a string specifying the label for y-axis

❖ Limits of x-axis (plt.xlim): list of two scalars

❖ Limits of y-axis (plt.ylim): list of two scalars

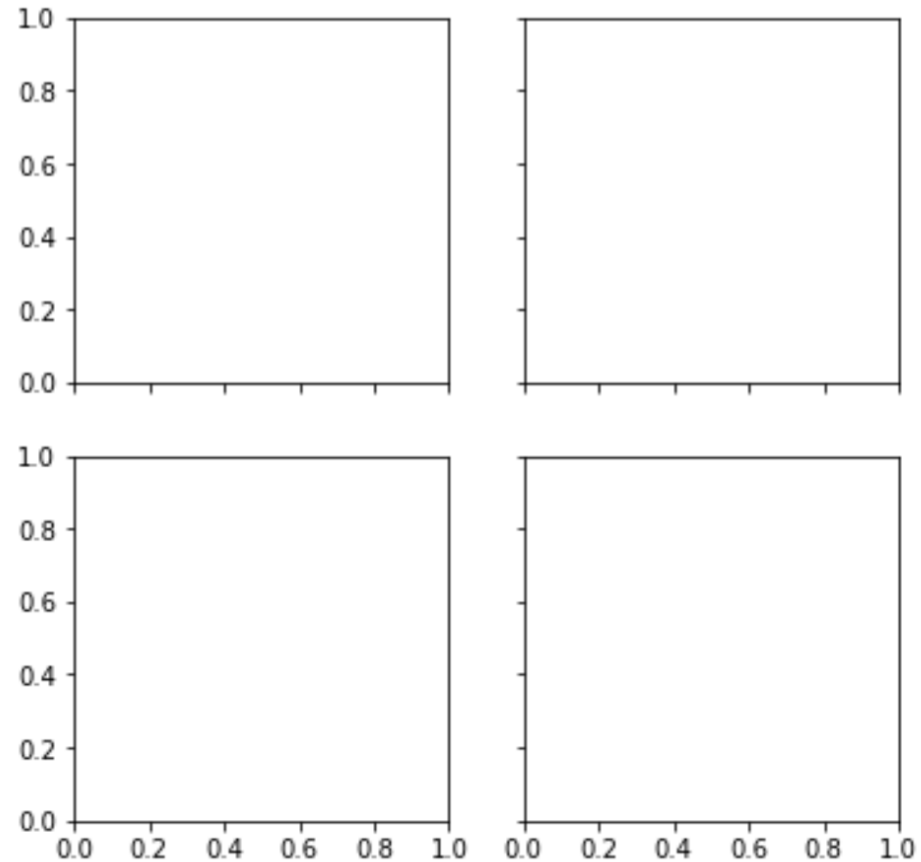❖ Setting the grids on (plt.grid): True or False

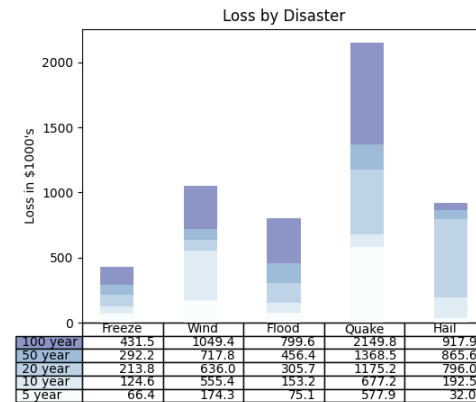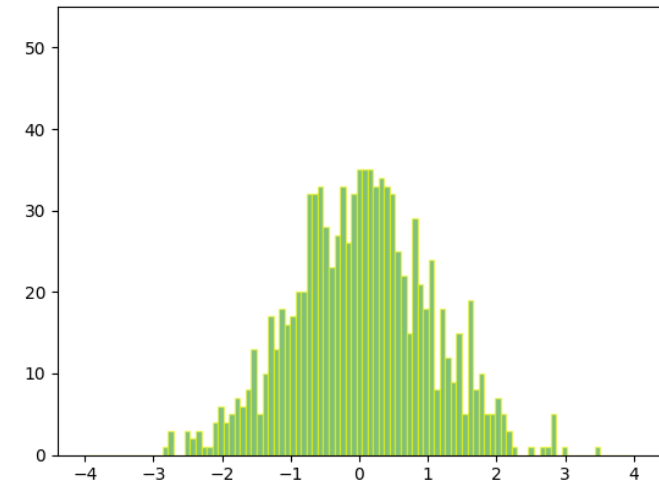❖ The legend (plt.legend)
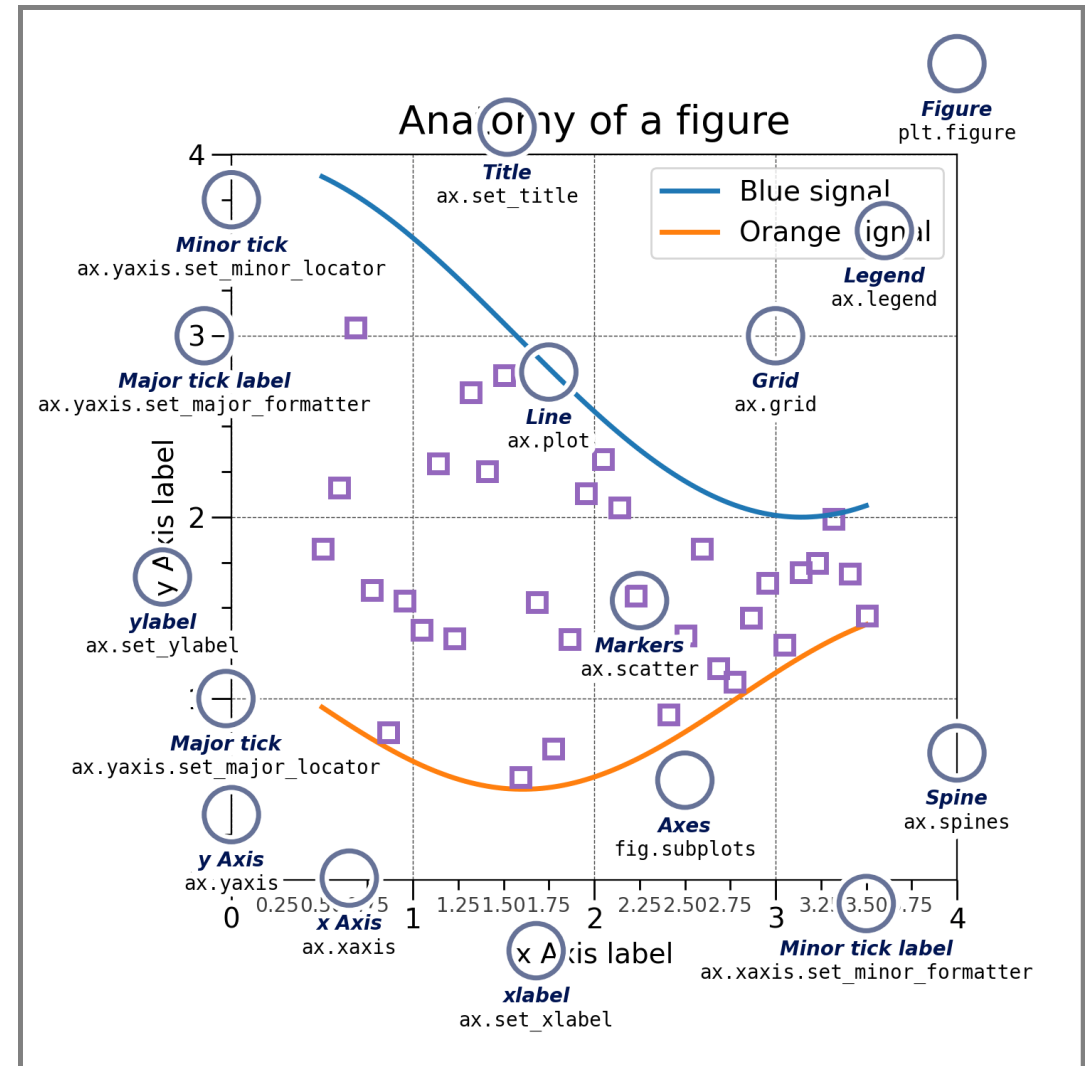
Questions?

Multiple Axes in the same Figure using Subplots

❖ Using the subplots command, we can create multiple axes within a figure.

```python
fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(6, 6),
                         sharex=True, sharey=True)
```

TILBURG ◆ UNIVERSITY

Useful Plots in Matplotlib

❖ Scatter plot

❖ Pie Chart

❖ Bar plot

❖ Histogram

❖ Boxplot
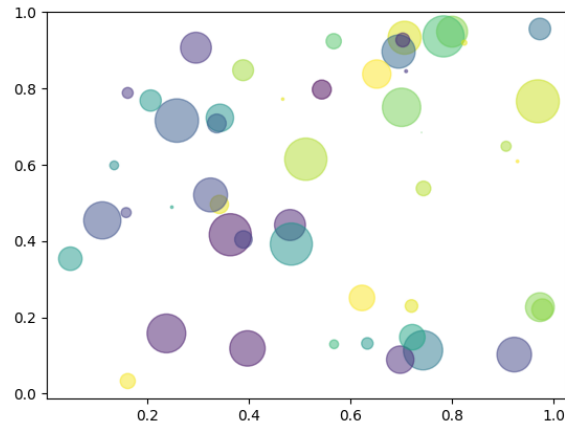
Scatter Plot

❖ Scatter plots are useful for:

    ❖ **Correlation Detection**: Scatter plots are ideal for detecting and showing the correlation between two variables.

    ❖ **Outlier Identification**: They are also useful for spotting outliers or anomalies in data.

    ❖ **Data Distribution Clarity**: Scatter plots can give a general idea of whether data points are tightly grouped or dispersed, which indicates the degree of variance within the data.
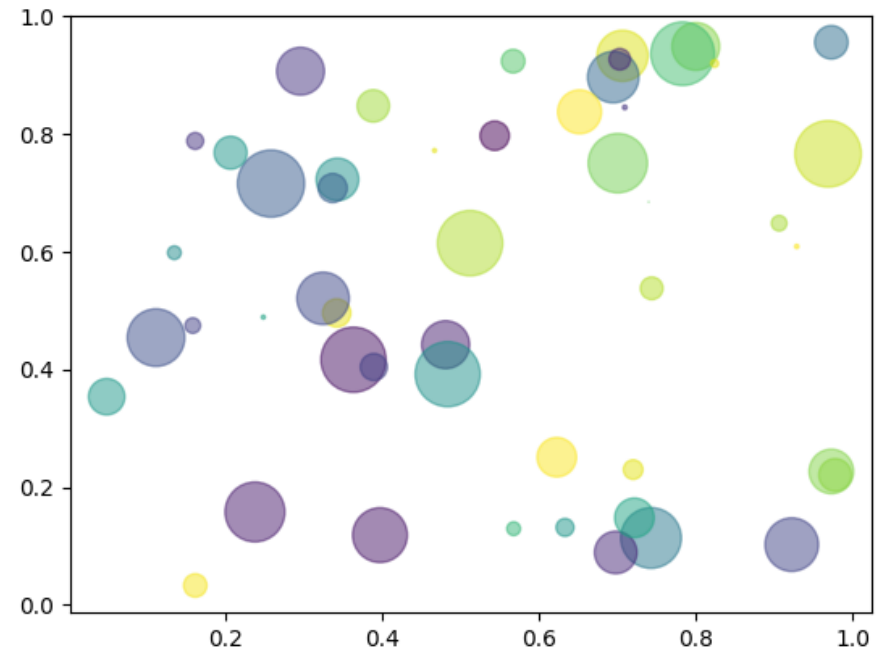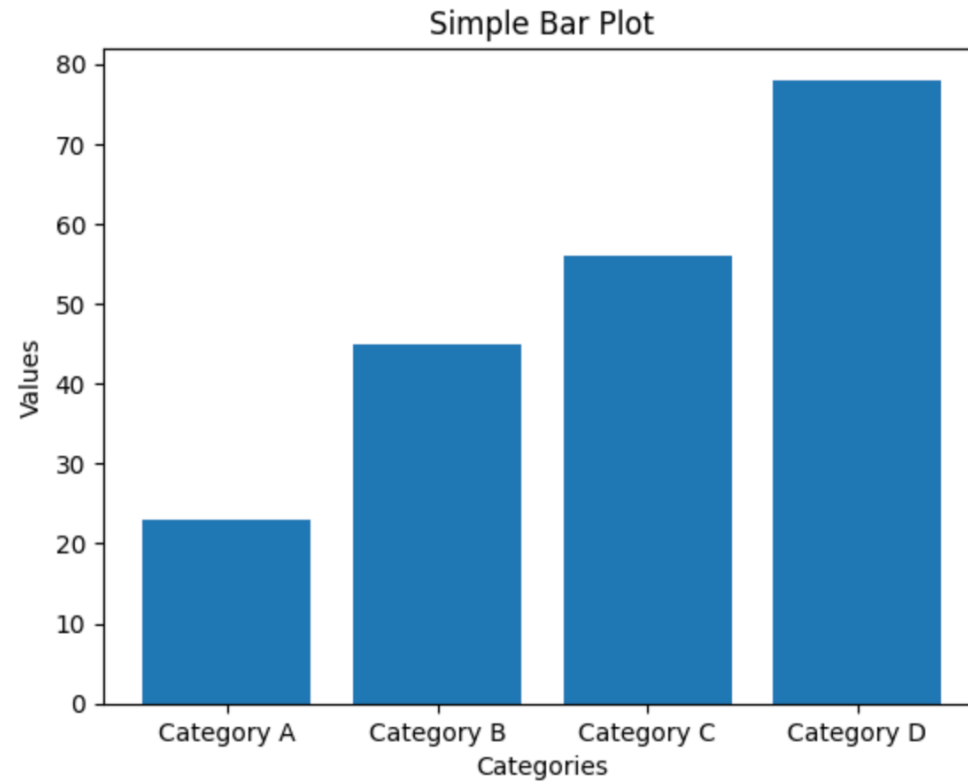
Pie Charts

❖ A pie chart is a circular graph where the circle is divided into sectors or slices.

❖ Pie charts are useful for displaying the percentage or the proportion of different categories in categorical data.

Budget Distribution Across Departments
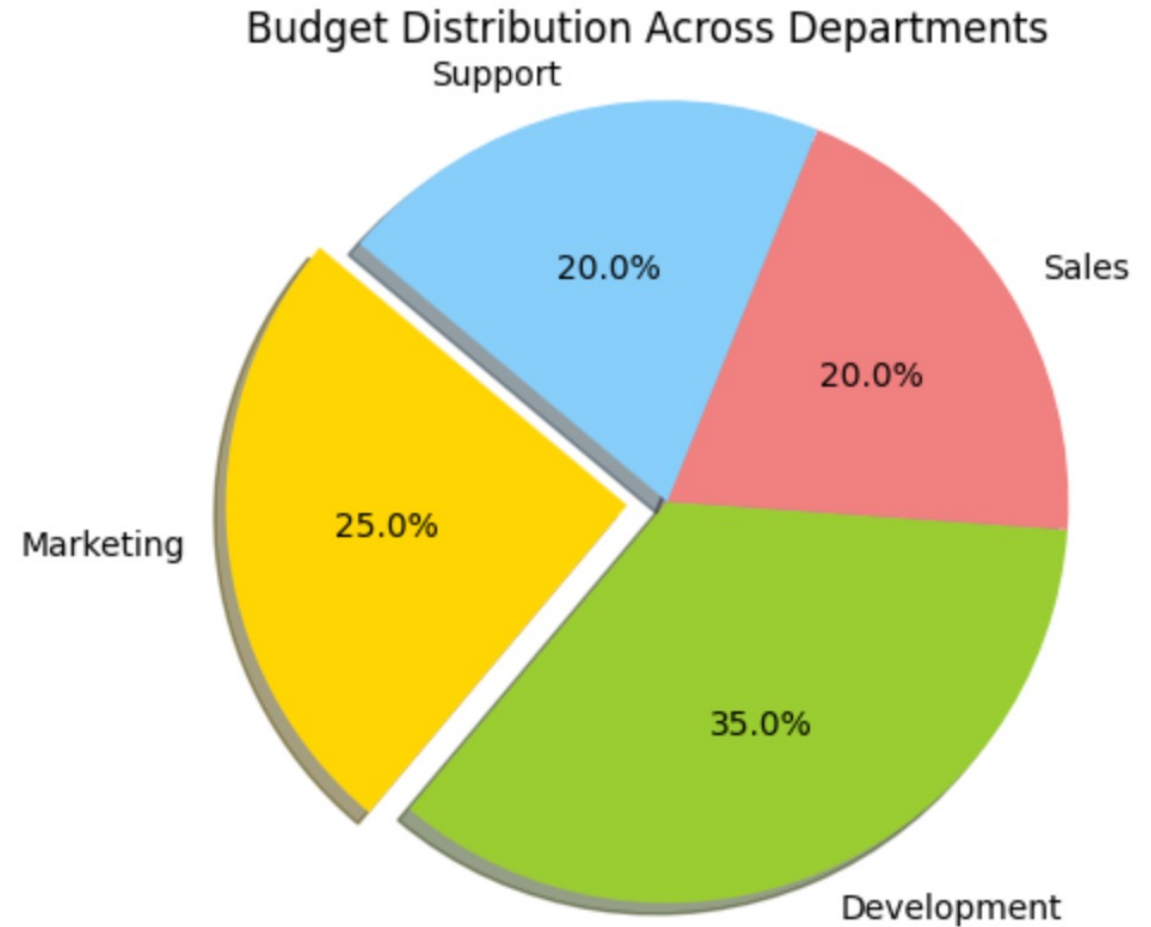
Bar Plots

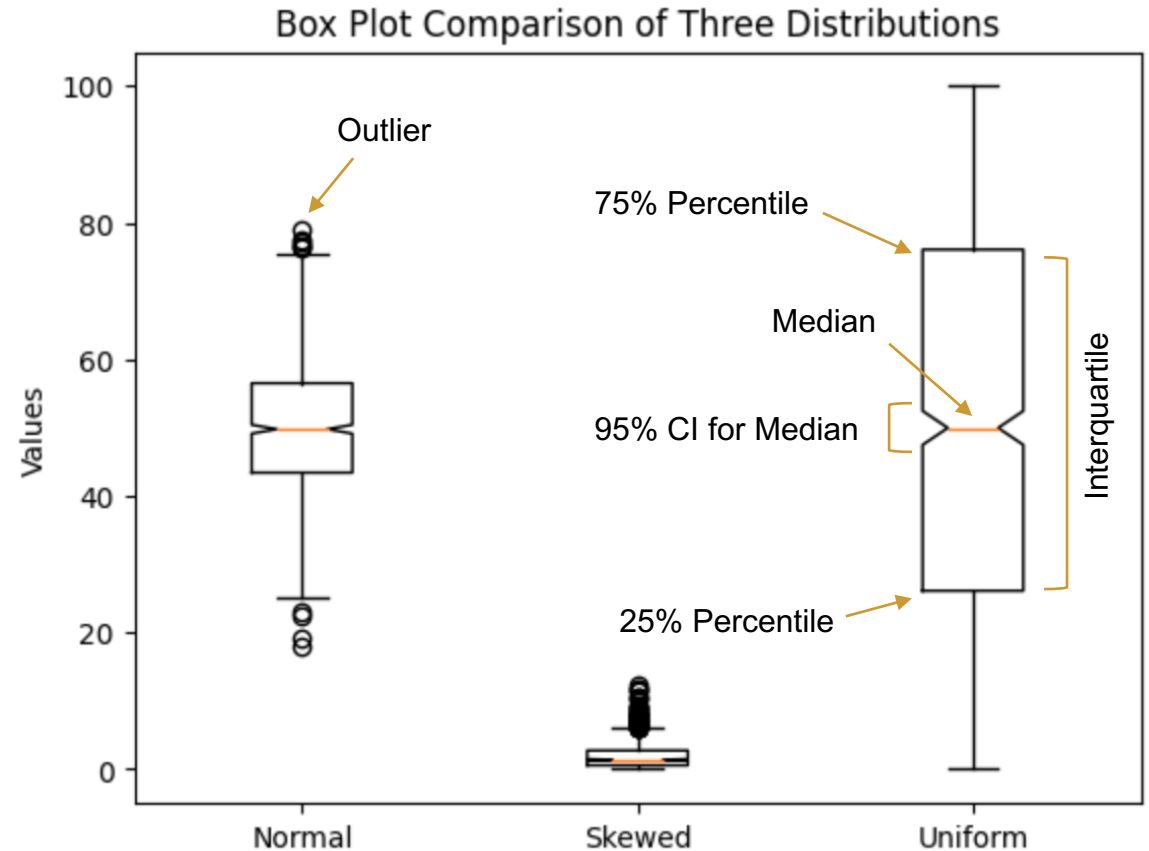❖ Bar plots are excellent for comparing different groups.

Histograms

❖ A histogram displays the frequency of data points within specified ranges (known as 'bins'). It is generally used for:

❖ **Data Distribution Insight:** Histograms are used to get a sense of the distribution of a numerical dataset.

❖ **Identifying Patterns:** They are useful for identifying patterns, outliers, and the spread of the data.

❖ **Comparing Data Sets:** Histograms can also be used to compare the distributions of multiple data sets.

TILBURG ◆ UNIVERSITY
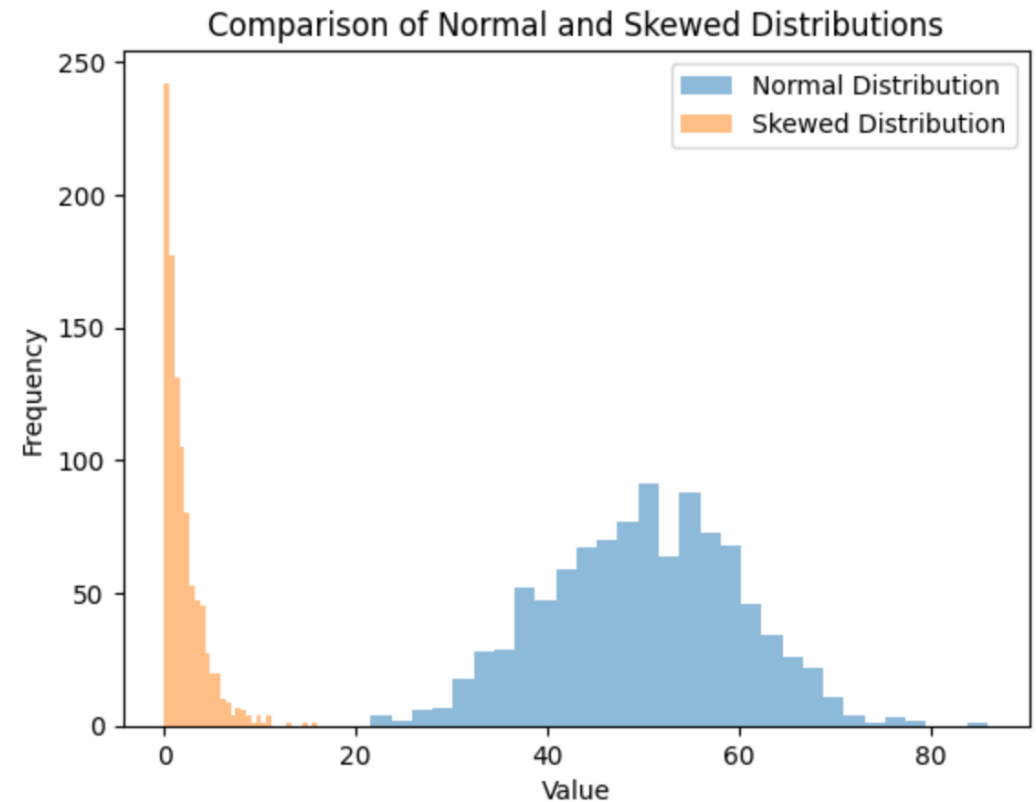
Boxplots

❖ Boxplot visually shows the distribution of numerical data and its skewness through displaying the data quartiles and averages. It is usually used for:

   ❖ **Outlier Identification**: Box plots are great for spotting outliers and data variability.

   ❖ **Comparison of Distributions**: They are useful for comparing the distribution of data across different categories.

   ❖ **Concise Data Representation**: Box plots can summarize a large amount of data in a visually digestible way.

Questions?

Thanks!