1. Scatter plot define

##1st - Scatter -used to scatter the values of x and y

1. Bar Plot

# Bar Plot -- Mainly bar plot is used to show the relationship between the numeric and categorical values. In a bar chart, we

## have one axis representing a particular category of the columns and another axis representing the values or count of the

## specific category. Bar charts are plotted both vertically and horizontally and are plotted using the following line of code:

1. Histograms plots

## Histograms - helps you to find out with respect to the numbers. (counts density on Y axis)

1. Pie charts:

# Pie Charts --A pie chart (or circular chart ) is used to show the percentage of the whole. Hence it is used when we want

## to compare the individual categories with the whole.

1. Area Chart

##Area Chart- Area chart or area plot is used to visualize the quantitative data graphically it is based on the line plot.

# 3-D Graphs-Now that you have seen some simple graphs it’s time to check some complex ones i.e 3-D graphs.

## Initially, Matplotlib was built for 2-dimensional graphs but later on, 3-D graphs were added to it.

## Let’s check how you can plot a 3-D graph in Matplotlib.

* **Machine learning** is like teaching computers to learn from data without explicitly programming them. It's a branch of artificial intelligence where computers analyse patterns in data to make predictions or decisions. So, instead of telling a computer exactly what to do, you give it examples and let it figure out how to do tasks on its own.
* The **mean** is just another way of saying "average." It's the sum of all the numbers in a set, divided by how many numbers there are. So if you add up a bunch of numbers and then divide by how many numbers there are, you get the mean. It's a way to find a typical value in a group of numbers.
* Sure! The **median** is like finding the middle number in a list of numbers. Imagine you lined up all the numbers from smallest to largest. The median is the one right in the middle. If there's an even number of numbers, it's the average of the two middle ones. It's a way to see what number is most typical in a set of data.
* The **mode** is the number that appears most frequently in a set of data. It's like finding the boss of the numbers because it's the one that shows up the most. If there are two or more numbers that appear with the same highest frequency, then the dataset is called bimodal or multimodal.
* **Variance** tells you how much the numbers in a dataset differ from the mean. Imagine you're looking at a bunch of numbers on a line. Variance measures how spread out those numbers are from each other. If they're all close to the mean, the variance is low. But if they're spread out, the variance is higher. It's a way to understand the "wiggliness" of the data.
* **Standard deviation** is a way to express how spread out the numbers in a dataset are. It's like a measure of how much the numbers tend to deviate from the mean (or average). If the standard deviation is small, it means the numbers are clustered closely around the average. But if it's large, it means the numbers are more spread out. Think of it as a measure of the typical distance between each number and the average.
* A **confidence interval** is like a range of values that we're pretty sure contains the true value of something we're trying to measure. It's like saying, "We're pretty confident that the real answer is somewhere between these two numbers." The confidence level tells us how confident we are that the true value falls within that range. So, the wider the confidence interval, the less precise our estimate, and the narrower it is, the more confident we are about our estimate.
* **Hypothesis testing** is a way for scientists and researchers to figure out if their ideas about how things work are right or not. It's like playing detective with data. Here's how it works:
* First, you start with a hypothesis, which is just a fancy way of saying an educated guess about something. Then, you collect some data to see if your guess is supported by evidence.
* Next, you analyze the data using statistical methods to see if there's enough evidence to support your hypothesis or if it's just a coincidence. If the evidence is strong enough, you can say your hypothesis is supported. If not, you might have to come up with a new guess and start the process over again.
* It's kind of like trying to solve a mystery using clues from the data.
* **Covariance** is a measure that tells us how two sets of numbers change together. Imagine you have two lists of numbers, and you want to see if when one goes up, the other tends to go up too, or if one goes down, the other tends to go down as well.

If the numbers tend to increase or decrease together, the covariance is positive. But if one set of numbers tends to go up while the other goes down, the covariance is negative.

In simpler terms, covariance is a way to see how closely related two sets of numbers are in terms of their changes.

**Correlation** tells us if two things are related or not. If they tend to change in the same direction (both go up or both go down), they're positively correlated. If one goes up while the other goes down, they're negatively correlated. And if there's no clear pattern, they're not correlated. It's like seeing how well two things match up or move together.

**The coefficient of variation** is a measure used to compare the variability of different sets of numbers, regardless of their units or scales. It's like a way to see how spread out numbers are relative to their average.

In simpler terms, it helps us understand how much the numbers in a dataset vary compared to their average value.

. How is the Mean Different from the Median? in simple language

Sure! Let's break it down:

The mean is the average of all the numbers in a dataset. You get it by adding up all the numbers and then dividing by how many numbers there are.

The median, on the other hand, is the middle number when the numbers are listed in order from smallest to largest. If there's an even number of numbers, it's the average of the two middle ones.

So, while the mean gives you the "center" of all the numbers based on their total value, the median gives you the "center" based on their position in the list.

**### 20. Why is Standard Deviation Preferred Over Variance?**

Standard deviation is often preferred over variance because it's in the same units as the original data, making it easier to interpret and compare. Variance gives us a measure of how spread out the numbers are, but the units are squared, which might not be as intuitive.

For example, if you're measuring the heights of students in centimeters, the variance would be in square centimeters, which isn't as easy to understand. However, standard deviation gives you a measure of spread in the same units as the original data (centimeters in this case), making it more straightforward to grasp and communicate.

### 21. What is Skewness?

Skewness is a measure that tells us how asymmetric or lopsided the distribution of numbers in a dataset is.

If the distribution is symmetrical (like a bell curve), it has zero skewness. But if it's skewed to the left, it means most of the numbers are on the right side and the tail stretches to the left. If it's skewed to the right, it means most of the numbers are on the left side and the tail stretches to the right.

In simpler terms, skewness helps us understand if one side of the dataset has more extreme values than the other**.**

### 22. What is Kurtosis?

Kurtosis is a measure that tells us how peaked or flat a distribution of numbers is compared to a normal distribution (like a bell curve).

If a distribution has positive kurtosis, it means it's more peaked and has heavier tails than a normal distribution. If it has negative kurtosis, it's flatter and has lighter tails.

In simpler terms, kurtosis helps us understand how much data is concentrated around the mean, and how extreme the tails of the distribution are.

### 24. What is the Interquartile Range (IQR)?

The Interquartile Range (IQR) is a measure of how spread out the middle half of a set of numbers is.

Here's how it works: First, you line up all the numbers from smallest to largest. Then, you find the median, which is the middle number. Next, you find the median of the numbers that are below the median (the lower half) and the median of the numbers that are above the median (the upper half).

Finally, you subtract the lower median from the upper median, and that's your Interquartile Range. It tells you how spread out the middle 50% of the numbers are.

In simpler terms, the IQR helps us understand the range of the middle part of the data, ignoring the extreme values at the ends.

### 25. What is a P-Value in Hypothesis Testing?

A p-value in hypothesis testing is like a report card for your hypothesis. It tells you how likely it is to get the results you observed if your hypothesis is actually true.

Here's how it works:

You start with a hypothesis, which is just a guess about something. Then, you collect some data to test if your guess is right or wrong.

After analyzing the data, you get a p-value. If the p-value is low (usually less than 0.05), it means that the results you observed are unlikely to happen just by chance if your hypothesis is true. So, you can feel more confident that your hypothesis might be correct.

But if the p-value is high (greater than 0.05), it means that the results you observed could easily happen by chance, even if your hypothesis is true. So, you might need to rethink your hypothesis.

In simpler terms, the p-value helps you decide if your hypothesis is supported by the evidence you collected or if it's just a lucky guess.

### 26. What is a Null Hypothesis?

Think of the null hypothesis as the "default" or "boring" hypothesis. It's a statement that says there's no effect or no difference between groups in your experiment or study.

For example, let's say you're testing a new medicine to see if it helps people feel better. The null hypothesis would be that the medicine has no effect, and any improvements you see in the people taking it are just due to chance.

Scientists usually try to disprove the null hypothesis because if they can show that there is an effect or a difference, it means their experiment or study has discovered something new or interesting.

In simple terms, the null hypothesis is like saying "nothing special is happening," and researchers try to prove it wrong.

### 27. What is the Central Limit Theorem?

The Central Limit Theorem is like a superhero of statistics. It says that no matter what kind of data you have, if you take enough samples and find the average of each sample, those averages will follow a predictable pattern.

Here's what it means:

Imagine you're measuring the heights of people in different cities. If you take a bunch of random samples from each city and find the average height of each sample, those averages will tend to form a bell-shaped curve, even if the original heights in each city didn't look like a bell curve at all.

In simpler terms, the Central Limit Theorem tells us that when we collect enough samples, the averages of those samples will tend to behave in a certain way, which makes it easier for us to make predictions and draw conclusions about the population as a whole.

### 28. How is Correlation Different from Causation?

Correlation and causation are like two friends who often get mistaken for each other, but they're actually quite different.

Correlation means there's a relationship between two things. For example, you might notice that as ice cream sales go up, so do sunglasses sales. That's a correlation—they're linked, but one doesn't cause the other.

Causation, on the other hand, means that one thing directly causes another. So, if you eat a lot of ice cream and then get a brain freeze, that's causation—the ice cream caused the brain freeze.

In simpler terms, correlation is just a connection between things, while causation means that one thing makes another thing happen. It's important to remember that just because two things are correlated doesn't mean one causes the other.

### 9. What is a Bar Plot?

A bar plot is a way to show data using rectangular bars, where the length of each bar represents the value of a particular category.

For example, imagine you're comparing the number of apples, oranges, and bananas sold at a fruit stand. You could use a bar plot to show the number of each fruit sold, with a separate bar for apples, oranges, and bananas. The height of each bar would represent how many of each fruit were sold.

In simpler terms, a bar plot helps you visualize data by using bars to show the values of different categories.

### 10. What is a Histogram?

A histogram is like a bar plot, but it's specifically used to show the distribution of numerical data.

Here's how it works:

You start with a range of numbers, called bins, and then count how many times each number in that range appears in your data. Each bin represents a range of values, and the height of each bar in the histogram shows how many data points fall into that range.

For example, if you're looking at the heights of students in a class, you might have bins for heights between 150-160cm, 160-170cm, and so on. The height of each bar in the histogram would then show how many students fall into each height range.

In simpler terms, a histogram helps you see how the values in your data are spread out across different ranges.

### 11. What is a Countplot?

A countplot is a type of bar plot that shows the number of occurrences of each category in a dataset.

For example, if you're counting the number of people with different hair colors in a group, a countplot would have a separate bar for each hair color, and the height of each bar would represent how many people have that hair color.

In simpler terms, a countplot is a visual way to see how many times each category appears in your data.