MIT | Academy of Engineering
(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

EDS → Makeup Assignment

Name :- Nikhil V. Jatale
Class :- CS8
Roll no :- 59
PRN NO :- 2024011200023

Q① explain Sample with example
Sample - require data set from huge dataset.

→ A sample is a subset of data selected from a larger population. It used in statistics to make inferences about the population without studying the entire group.
Example :-
A college has 1000 students. Surveying all students is difficult, so a researcher selects 100 students randomly. This group of 100 is a sample representing the whole population.

Q② explain population with example
population - complete - full - raw data set.

→ A population is statics refers to the complete set of data or all possible observations of a particular group.
Example :-
If a college has 1000 students,

the population is all 1000 students. If you survey all of them, you are studying the population.

Q3] Explain the Statistical Analysis (SA) of data

Definition :-
Statistical Analysis (SA) involves collecting, reviewing, and interpreting data to identify patterns, trends, and relationships. It helps in decision-making and forecasting.

Types of SA :-

(1) Descriptive statistics
↳ mean, median, mode, Standard deviation.

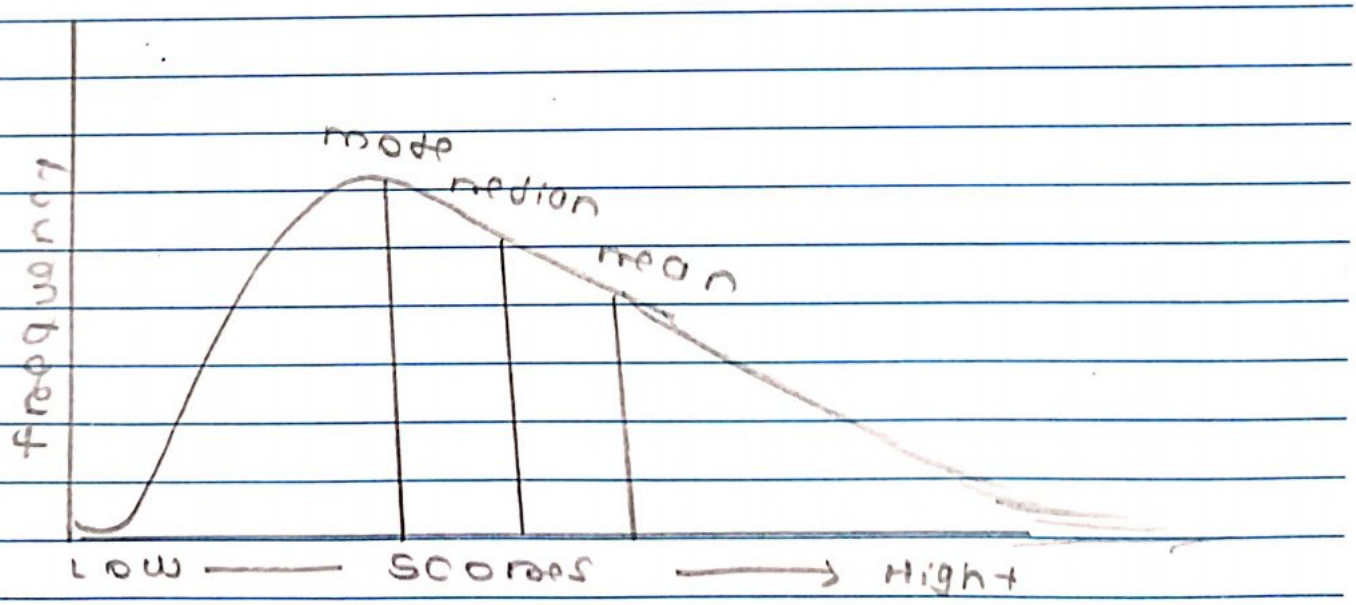
(2) Inferential statistics
↳ Hypothesis testing, confidence intervals.

Q4] Explain Normal Distribution, Right Skewed Distribution, left Skewed Distribution with example and graph.

⇒

(ii) Right-Skewed Distribution :-

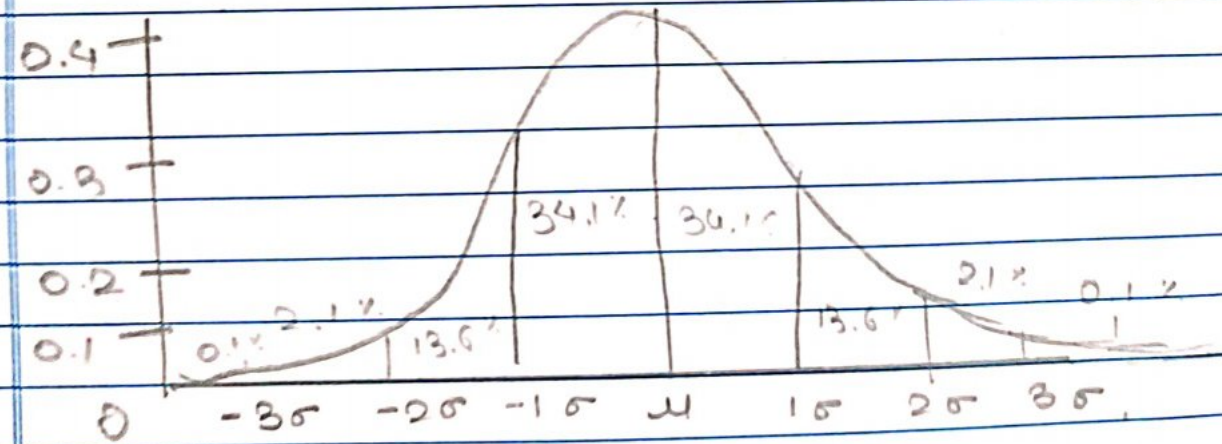
- Shape : Asymmetrical with a longer tail on the right side.
- mean, median, mode :- mean is greater than the median, which is greater than the mode.
- Skewness : Positive.
- Example :- Income distribution is often right-skewed, as a few individuals have very high incomes, while most have lower incomes.
- Graph : A curve with the peak to the left of the center, and the right tail extending further than the left tail.



⇒

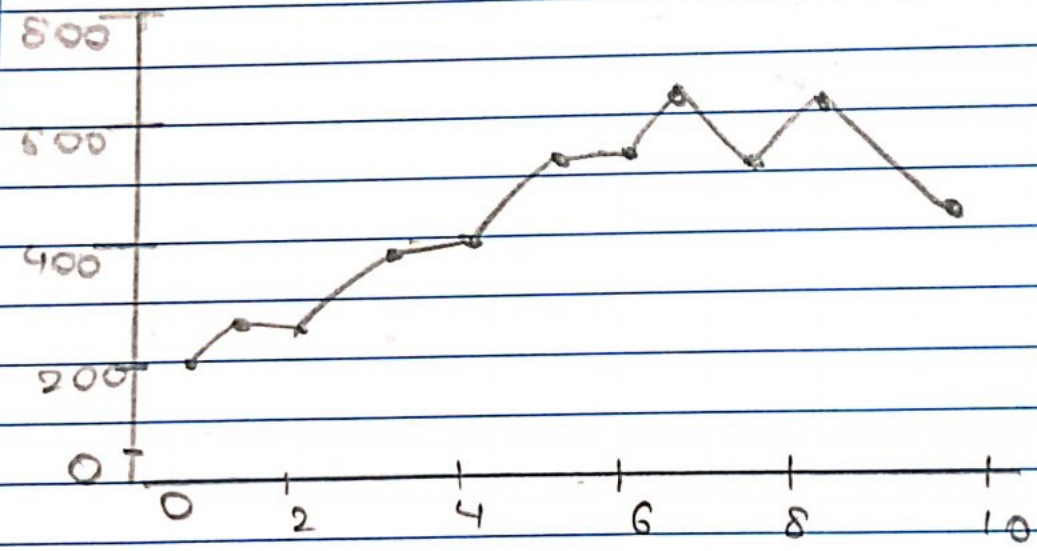
(i) Normal Distribution :-

- Definition :- A symmetrical distribution where data is clustered around the mean and the curve gradually tapers off on either side.
- Shape :- Bell-shaped with the peak at the center, representing the mean.
- Example :- Heights of adults in a population, where most heights cluster around the average height, and fewer people are very tall or very short.
- Graph :- A symmetrical curve with the mean at the center.



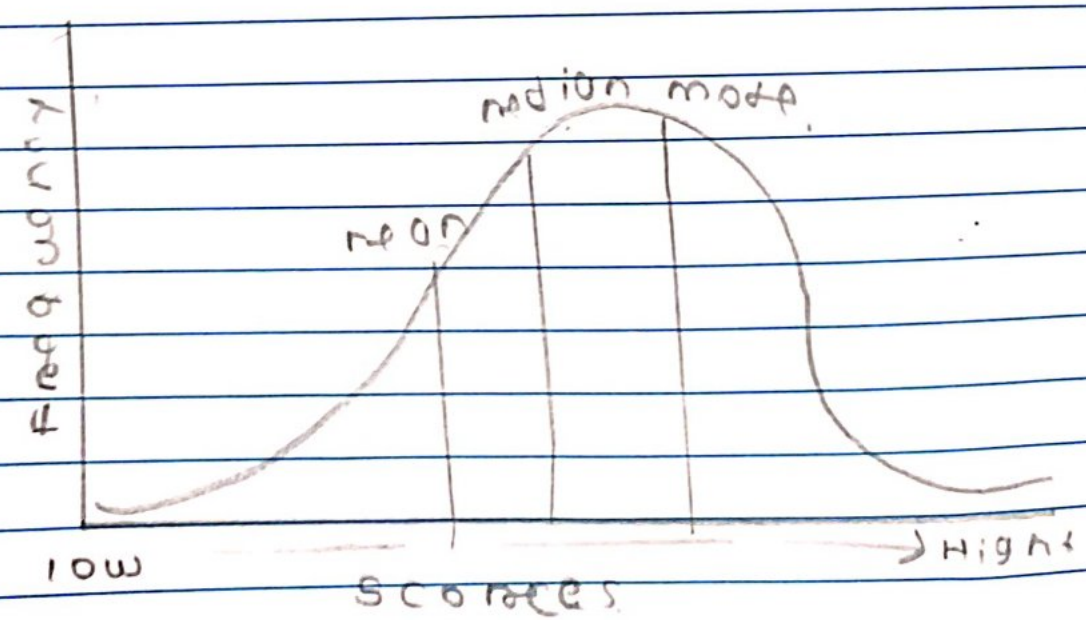
Q5 Explain different visualization mode with example and graph / design.

1] Line chart :- A line chart is used to display trends over time. It shows the relationship betn two variables where one variable is typically time. For example, a line chart could show the sales performance of a company over the past five years.

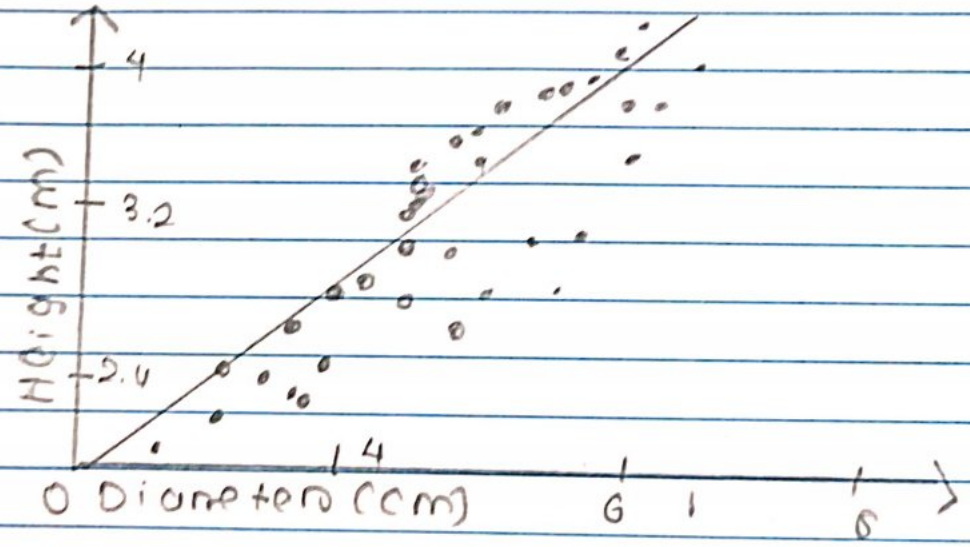


(iii) Left - Skewed Distribution:-

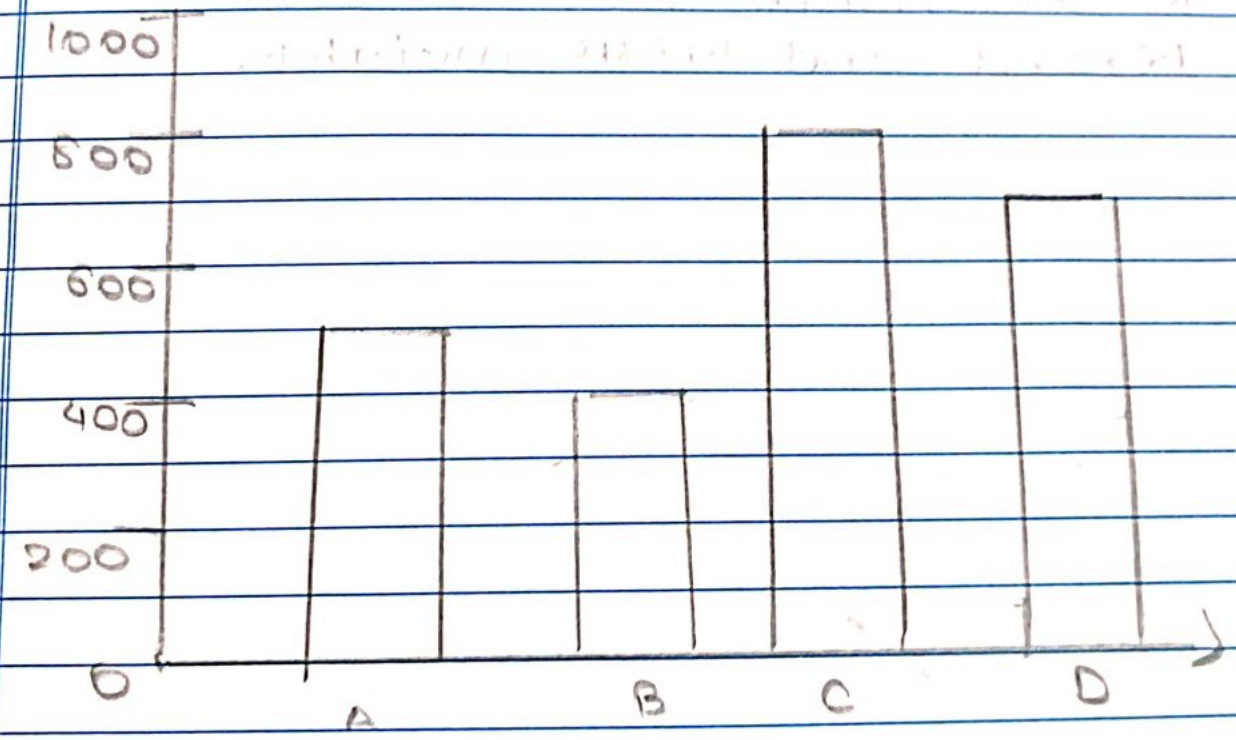
- Shape :- Asymmetrical with a longer tail on the left side.
- mean, median, mode :- mean is less than the median, which is less than the mode.
- Skewness :- negative
- Example :- Exam scores in a class can be left-skewed if most students score high and a few score very low.
- Graph : A curve with the peak to the right of the center, and the left tail extending further than the right tail.



(iii) Scatter plot :- A scatter plot is used to represent the relationship between two variables. Each point on the plot represents the value of both variables for a single observation. This allows us to visualize trends and potential correlations betn the two variables. For example, a scatter plot could show the relationship betn a person's height and their weight.



2] Bar chart :- A bar chart is used to display categorical data. It uses rectangular bars to represent the magnitude or frequency of each category. For example, a bar chart could be used to compare the number of students enrolled in different departments at a university.

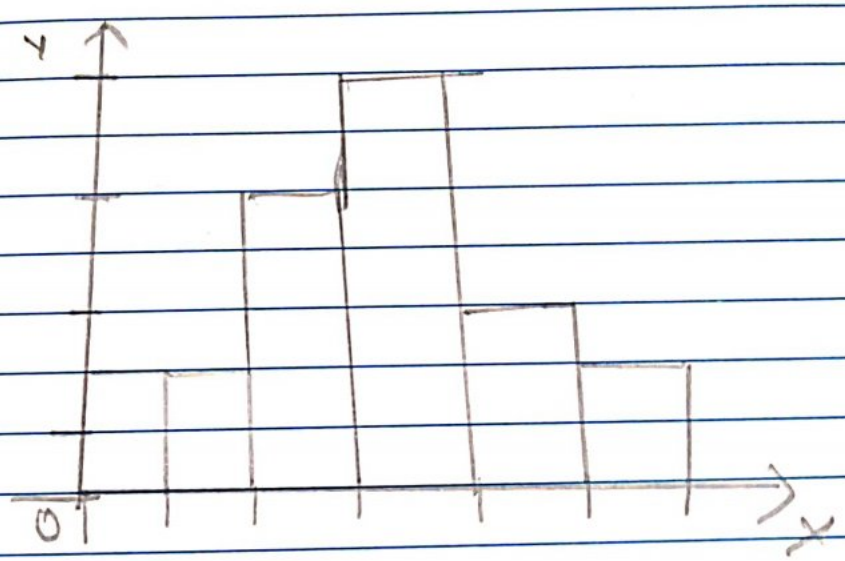


5] Pie chart :-

A pie chart is used to display the proportions of a whole. Each slice of the pie represents a different category, and the size of the slice is proportional to the percentage of the whole that the category represents. For example, a pie chart could be used to show the distribution of market share among different companies in an industry.

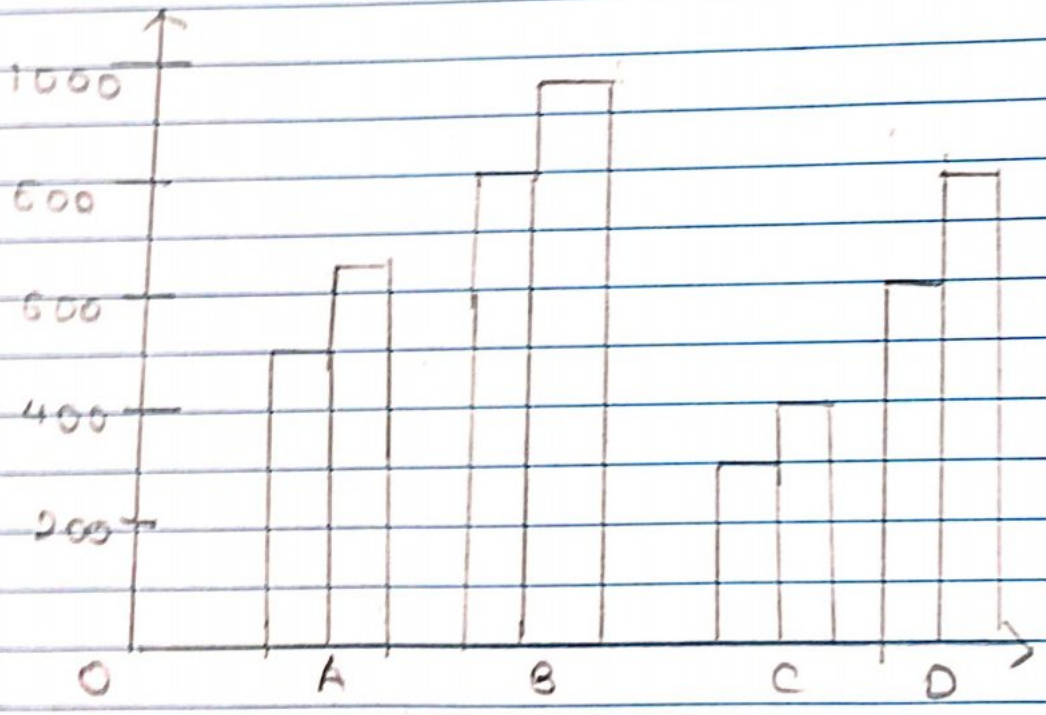
4) Histogram plot :-

A histogram plot is used to show the distribution of a single variable. It divides the range of values into bins and then shows the frequency of observations falling into each bin. This helps us understand the shape and spread of the data. For example, a histogram could be used to show the distribution of exam scores in a class.



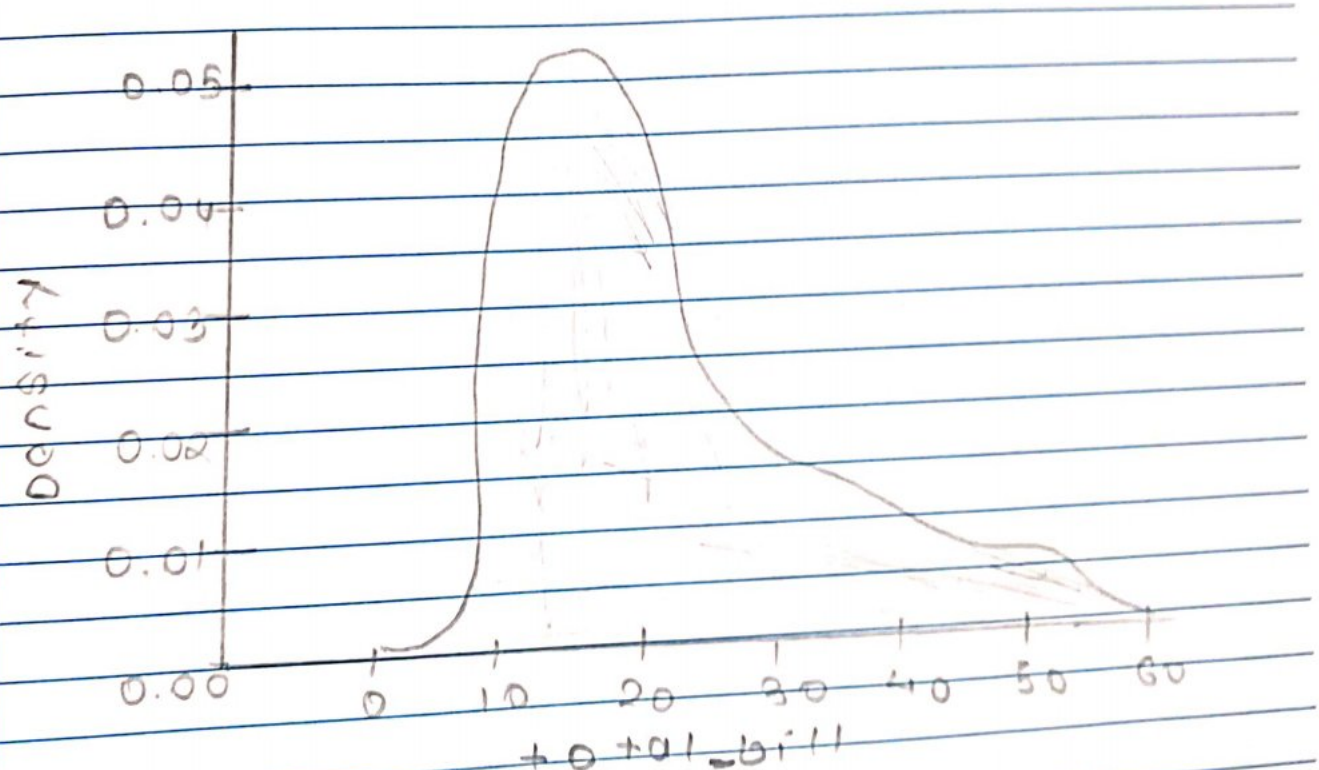
7] Group plot :-

A group plot is used to compare different datasets. It often involves comparing the same type of data for different groups or categories. For example, a group plot could be used to compare the performance of different products in the same market.



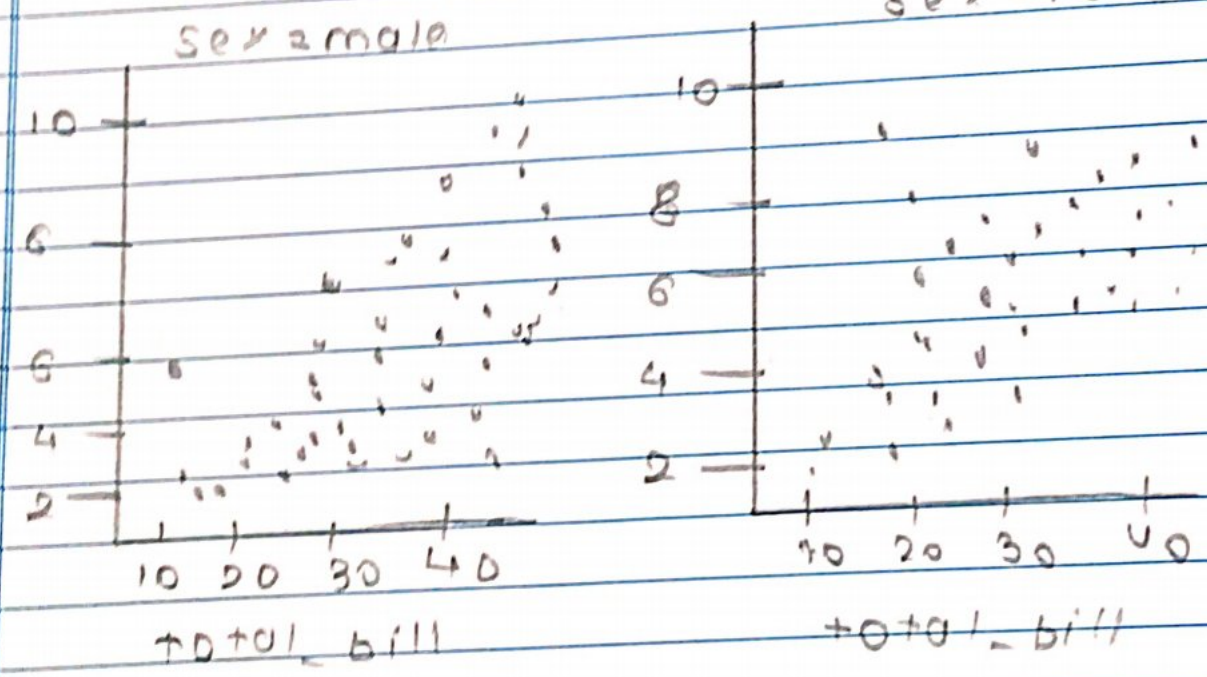
6] Density Plot:-

A density plot is a smoothed version of a histogram. It estimates the probability density function of a variable. This can be useful for visualizing the distribution of a variable when we want a smoother representation than a histogram. For example, a density plot could be used to show the distribution of a population's age.



9] Facet Grid plot :-

A facet grid plot creates multi-plot grids for conditional relationships. It divides the data into subsets based on one or more variables and then displays a plot for each subset in a grid format. This allows us to explore how the relationship between two variables changes under different conditions.



8] Panels :-

Panels refer to multiple plots arranged in a grid. This allows us to easily compare and contrast different plots and datasets. for example we might have a panel of line charts showing the performance of different companies over time.