



Notes

5

DATA INTERPRETATION AND STATISTICAL DIAGRAMS

In our daily life we come across different types of information through newspaper, television, computers and conversation. Some of this information is quantitative and others are qualitative, which can not be measured numerically. Statistics is concerned mainly with information that is quantitative or measured numerically. In this lesson we will study about statistical data and their representation through various types of diagrams and maps.



OBJECTIVES

After studying this lesson, you will be able to:

- distinguish between statistical information (statistical table) and diagram;
- calculate mean, median, mode and percentile;
- recognise various types of diagrams such as a line, bar, pie and star diagram and a dot map;
- construct line, bar, pie and star diagrams;
- select a suitable diagram for the given data;
- explain merits and demerits of each diagram.

5.1 STATISTICAL DATA : PRESENTATION AND INTERPRETATION

Let us take a poor man from the United States of America (USA) and a rich man from India. If the income of the Indian is higher than the income of the poor man of the USA, can we say that India is richer than the USA? Certainly not. Why? Because comparison is between two specific persons from the USA and India, which does not represent their countries so far as their individual income is



Notes

concerned. For any such comparison we have to see the income of larger population in the USA as well as in India. For this we will have to collect information about the annual income of individuals, agricultural production, industrial production, unemployment rates, total population of different areas etc. All such information will be numerical and will relate to a large number of individuals or areas. Numerical information related to the measurement of groups or masses is termed as data, (singular is datum). When information is related to an individual or event, it is not data.

In geography, statistical data play a very important role. The data gives us numerical information about geographical facts such as temperature, rainfall, agriculture/ industrial production, population, etc. We arrange them, analyse them, draw valid conclusions from them.

Source of Data

There are two sources from where data can be obtained. They are:

1. Primary source and
 2. Secondary source
1. **Primary source:** If the data obtained through field investigation is called primary data. It is a lengthy process requiring a lot of time, money and manpower.
 2. **Secondary source:** If the data are the published in the form of reports and tables prepared by various public agencies for general use is called secondary data. For the user the data from secondary sources are thus less expensive, time saving and do not require large manpower for their collection. Secondary data, however, is very general. Census of India is one of the best sources of secondary data.

5.2 PRESENTATION OF STATISTICAL DATA

The data collected through various sources needs to be processed statistically for precise explanations. Very often it becomes necessary to obtain a single representative value for the whole data set. The statistical measures that enable us to work out a single representative figure for the entire data distribution, is known as central tendency. Measures of central tendency help us to compare different distributions besides being representative for each distribution. These measures normally denote the central points of values, distance and occurrence in a distribution. The commonly used measures of central tendency are:

- (i) Arithmetic mean or average
- (ii) Median
- (iii) Mode
- (iv) Percentile



(i) Arithmetic Mean

It is most frequently used and is calculated by adding the sum of all individual values in a distribution and dividing the sum by the total number of individuals. For example, the production of rice per acre in five districts is 10, 8, 12, 9 and 6 quintals. The average production of rice for these districts is :

$$\frac{10 + 8 + 12 + 9 + 6}{5} = \frac{45}{5} = 9 \text{ quintals per acre}$$

The arithmetic mean is expressed in the form of equation noted below:

$$\bar{X} = \frac{\sum X}{N} \quad \dots(i)$$

Where \bar{X} is the mean value,

$\sum X$ = The total of X values

N= Number of individuals/observations.

The arithmetic mean can be easily worked for small ungrouped data. However, when the number of observations are large and data is in the form of frequency distribution of groups, arithmetic mean will be worked out with the help of following equation.

$$\bar{X} = \frac{\sum fm}{\sum f} \quad \dots(ii)$$

Where \bar{X} is the arithmetic mean,
f is the frequency,
m is the mid value of the classes

Example

Calculate the arithmetic mean from the temperature (in degree celsius) data given in the following table.

Table 5.1

Temperature Classes <i>x</i>	No. of days <i>f</i>	Mid values <i>m</i>	<i>fm</i>
1-05	20	3	60
06-10	24	8	192
11-15	44	13	572
16-20	72	18	1296
21-25	76	23	1748
26-30	60	28	1680
31-35	52	33	1716
36-40	4	38	152
41-45	8	43	344
	$\sum f = 360$ days		$\sum fm = 7760$

Notes

1 quintal = 100 kg



Notes

From the above

$$\Sigma fm = 7760$$

$$\Sigma f = 360$$

$$\bar{X} = \frac{7760}{360} = 21.56^{\circ}\text{C temperature}$$

Merits of the Arithmetic Mean

1. It is easy to understand the complete idea of the distribution and simple to workout.
2. It is the average of the values in a distribution. Hence, it has a balancing property in case of sample surveys. Therefore, the mean is the centre of gravity.
3. It is widely used in case of normal distributions.

The arithmetic mean has certain limitations. It is affected by the extreme values especially when they are large. For example, income variations are very wide in case of Indian population.

(ii) Median

Median is the middle most positional average. It is worked out by arranging data in an ascending or descending order. For example, the value of the median is worked out by adding 1 to the number of observation and the sum divided by two. It is expressed as:

$$\text{Med} = \frac{N+1}{2} \quad \dots(\text{iii})$$

For example if we are interested in working out the median latitude and longitude for the country, we must arrange these distributions in a tabular form.

Latitudinal Extent of the Mainland of India (8' 4' N to 37' 6' N)

Table 5.2

8	9	10	11	12	13	14	15	16	17
18	19	20	21	22	23	24	25	26	27
28	29	30	31	32	33	34	35	36	37

The median or middle most latitude of India is 23°N which is close to the Tropic of cancer (23° 30' N). Since mainland of India starts from 8° 4' N which is a part of 9th latitude and extends up to 37° 6' N which covers the 37° latitude completely,



Notes

hence the latitudinal coverage of India is approximately 29° latitudes. The median latitude is therefore, 23°N i.e.

$$\text{Med} = \frac{N + 1}{2} = \frac{29 + 1}{2} = \frac{30}{2} = 15^\circ + 8^\circ = 23^\circ\text{N}$$

8° (Southern tip of India) + 15° (median value) = 23° (middle east latitude of India). Similarly, we can also work out the median value for the longitudinal extent of India. The Longitudinal Extent of India ranges between = 68°7' E to 97°25'E.

The median or middle most longitude for the country is 83°E.

Table 5.3

68	69	70	71	72	73	74	75	76	77
78	79	80	81	82	83	84	85	86	87
88	89	90	91	92	93	94	95	96	97

Longitudes are used to calculate local time, standard time of a nation and international time, which is linked to Greenwich Mean Time (GMT). Indian Standard Time is calculated keeping 82°30' E longitude as the base. The median longitude for the country is 83°E which is close to the standard meridian used for Indian Standard Time calculation.

$$\text{Med} = \frac{N + 1}{2} = \frac{29 + 1}{2} = \frac{30}{2} = 15^\circ + 68^\circ = 83^\circ$$

Merits of Median

1. Being the middle most value, median remains unaffected by the extreme values in the distribution as in the case of arithmetic mean.
2. It is a partition value which divides the series into two nearly equal parts and remains the centre of gravity.
3. However, it cannot be worked out without putting data in an ascending or descending order. If data are large, it might be a time consuming and tedious job. The values of median will be erratic if one or two items are added or subtracted from the series.

(iii) Mode

It is one of the important measures of central tendency. The maximum concentration of items occurring in a distribution is known as mode. The value which occurs most frequently is identified as mode in case of ungrouped data. Similarly, for grouped data the mode can be calculated by identifying the class with the highest frequency. The mode denotes the centrality of the occurrence of an item in the distribution. The distribution of rural settlements in Uttar Pradesh is given below. Work out the mode for the data.



Notes

Table 5.4
Distribution of Rural settlements in Uttar Pradesh (2001)

Size of Rural Settlements	Very small (Below 500 Population)	Small (500-999)	Medium (1000-1999)	Large (2000-4999)	Very Large (5000 and above)
Proportion of distribution	16.70	23.45	47.96	10.60	1.29

Solution: Arrange the data in a sequence (either from small to large or from large to small). Put up the frequency values against each. Now compare the frequencies. The distribution registering maximum frequency in medium size of rural settlement (1000-1999) 47.96 is identified as ‘mode’.

Merits of the Mode

1. It is the most typical value of a series. Mode can be located easily by the inspection and can be used by common people also.
2. The occurrence of a few extreme values does not affect the mode, since it is the most typical value of series.

It is, however, not a significant measure of central tendency unless the number of observations is large. Both in case of uniform as well as skewed distributions, mode ceases to be a measure of central tendency.

(iv) Percentiles

Percentile is a measure which divides a series into 100 equal parts. It helps to understand various classes or categories that constitute a distribution. It is expressed as:

$$P = \frac{P \times N}{100} \quad \dots (iv)$$

Where P is the percentile and N is the number of observations.

There are 99 percentiles, P_1, P_2, \dots, P_{99}

Table 5.5
Distribution of Monthly Income Among Households of a locality

Income group (Rs.)	Actual Number	Percentage Distribution
Economically weaker sections (Below Rs.500)	112	56.0
Lower Income Group (500-999)	41	20.5

Middle Income Group (1000-4999)	29	14.5
High Income Group (5000 and above)	18	9.0
Total	200	100.00

Table 5.6
Distribution of Per Capita Monthly Income of the
Households of a locality

Income group (Rs.)	No. of Frequency Households	Cumulative Frequency
Below 500	112	112
500-999	41	153
1000-4999	29	182
5000 and above	18	200
Total	200	

Let us calculate 60th percentile as P_{60} .

Now $P_{60} = 60 \times 200 \div 100 = 120$

The income lies in the group 500–999.

5.3 REPRESENTATION OF STATISTICAL DATA THROUGH DIAGRAMS

The data collected either through primary source or through secondary source, are raw and unorganized form. They do not give a clear picture because some values are very large, some are low and some are in between low and high. These are all scattered values lying here and there. The data becomes clear once they are organized and put in to some systematic tabular form. Statistical tables are very handy and represent the data in a systematic and manageable form. A direct-mental comparison of such values is also possible if these are represented through diagrams.

WHY DO WE REPRESENT DATA THROUGH DIAGRAM

The below mentioned points reveals the advantages of diagrams over the raw data. The advantages are as follows:

- (i) The diagram creates greater interest in the subject matter which has been represented by it.



Notes



Notes

- (ii) It clarifies and simplifies the subject matter.
- (iii) They help in making quick and accurate comparison of data.
- (iv) They bring out hidden facts and relationship and can stimulate analytical thinking.
- (v) It is more illustrative and attractive than the statistical information

Following are important diagrams/maps through which various types of statistical data can be represented:

- (i) Line graph
- (ii) Bar diagram
- (iii) Pie diagram
- (iv) Star diagram

(i) LINE GRAPH

There are certain variables whose values fluctuate with time, like temperature of an area or rainfall etc. There are some other variables which increase or decrease with time like population, agricultural and industrial production and prices of various commodities. The data for all such variables are collected and tabulated with reference to time. If we plot such data on a graph paper in such a way that time is plotted on x-axis and values of the variables are plotted on y-axis and join the points by straight lines what we get is known as a line graph.

Example

Average monthly maximum temperature of a place 'A' is given below for 12 months. Plot the data by a line graph.

Table 5.7

Month	Jan	Feb	Mar	Apr	May	June	July	August	Sep	Oct	Nov	Dec
Temp. °C	24.5	26.6	32.2	38.1	42.5	44.3	40.4	33.4	30.2	29.7	29.2	25.0

In order to plot the given data, the most suitable diagram would be a line graph since the values of the temperature are given against months. Plot months on the x-axis and temperature on the y-axis. Keeping the temperature of various month, plot 12 points. If we join these points by straight lines as given in figure 5.1 we get the required line graph.

We observe from the table that the monthly average of maximum temperature is least in the month of January and it rises marginally in February. In March and April it increases rapidly and reaches maximum of 44.3°C in June. It remains quite hot in July also. From August onward the temperature starts declining again.



Notes

This conclusion, about the temporal variations or changes in temperature does not require any elaborate description in the presence of the line graph. Anybody looking at the linegraph given in figure 5.1 can quickly make out the picture of this cyclic variations in the temperature of the area.

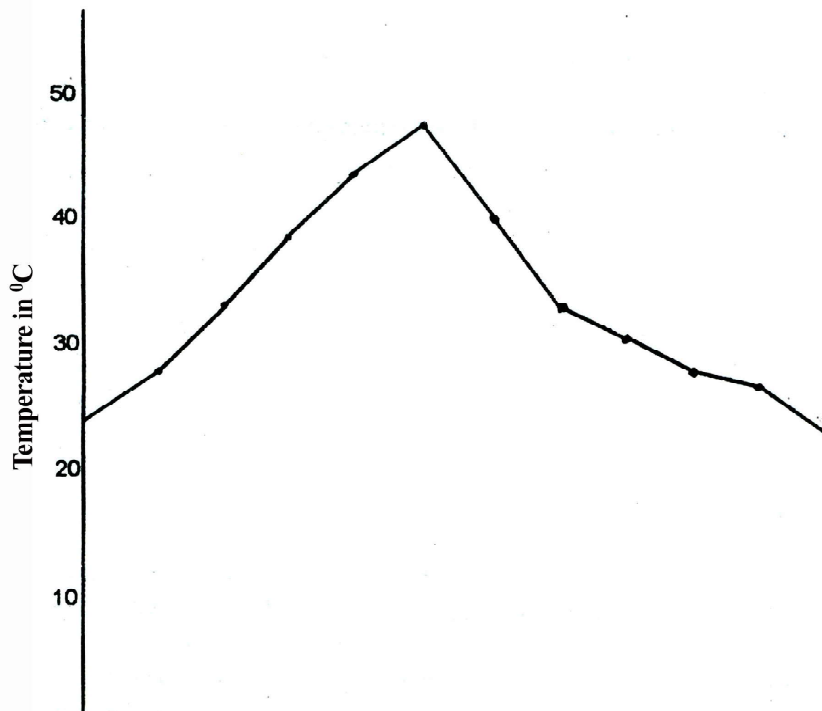


Fig. 5.1 Line Graph

Rules for making line-graph

1. Time which is an independent variable is plotted on x-axis and temperature, rainfall, production or any other variable which is varying over time is plotted on y-axis.
2. All such points obtained by plotting the values on x-axis and y-axis should be jointed with the help of straight lines.
3. Since units of both x-axis and y-axis are different, separate scales should be chosen for these. For x-axis, the time unit-could be hours, days, months, years or any other unit of time. For y-axis the unit-could be °c, cm, tonnes or any other unit.
4. Normally vertical scale should start with zero so that the absolute magnitude of the values are represented. However, if most of the variations start after some fixed value, that value may start from origin of y-axis, for example take the values 12050, 12020,12180,12200,12140, 12040, 12120.....etc. As among these values variations are found after 12000, we can take 12000 as the starting value on the y-axis.



Notes

- In a line graph time variable in most of the cases, is at fixed intervals. It could be hours, days, months or years etc. Although this rule is not a necessary condition but is generally observed for the sake of clarity.

Characteristics of a line graph

- It shows the past as well as the present trends in the variations of a phenomenon.
- With the help of it the intermediate values can be estimated (interpolated) as well as future values can also be predicted (extrapolated).
- In comparison to mathematical relationship between time and the values of a variable, the relationship shown through line graph is only approximate.
- It occupies more space as compared to a mathematical relationship.

Compound line graph

Some times more than one variable can also be plotted on a line graph to compare their relative changes. For example, we can plot export and import figures on one graph for several years. It will give the changes in export in relation to imports. The difference between the two, will also give trade deficit. Similarly we can plot birth rates and death rates of a country for several years. The difference between the two in this case will give natural growth rate of the population. We may also plot the production of various agricultural crops to see as to which crop is maintaining upward trend and which is going downward. We can also plot the monthly maximum and minimum temperature on the same graph. Difference between the two will give monthly range of temperature.

Example

Line graph showing temporal variations in more than one variable is known as compound line graph.

The estimated Crude Birth Rate (CBR) and Crude Death Rate (CDR) of India are given below for the following years. Plot the data on a line graph.

Table 5.8

Year	1921	1931	1941	1951	1961	1971	1981	1991
CBR (per 1000)	49	47	45	43	44	42	37	30
CDR (per 1000)	49	37	33	31	26	20	15	10

The data is plotted on the graph given in figure 5.2. There are two types of lines representing two different types of population data as indicated in the table. The compound line graph of crude birth rates and crude death rates shows a decline in both during 1921-1991. However, the graph shows a steep fall in crude death



Notes

rate in relation to crude birth rate. This increasing gap is also highlighted on the graph by the shaded area between CBR and CDR.

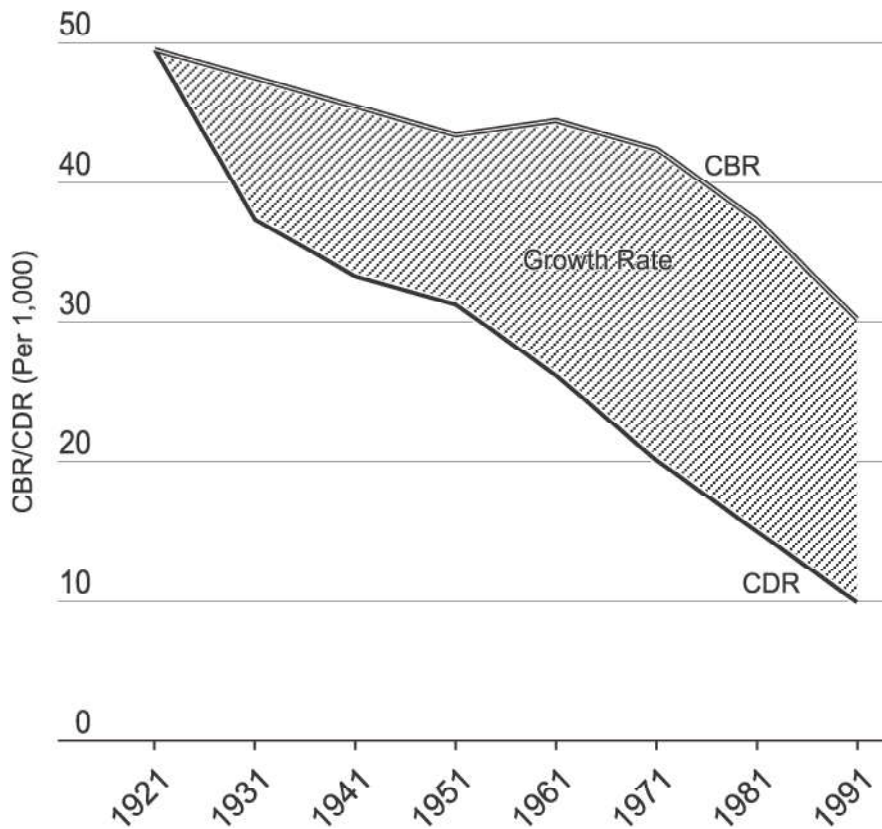


Fig. 5.2 Compound Line Graph

(ii) BAR DIAGRAM

Some times the values of a variable are given for areas, commodities or for anything other than time. In such cases these values are represented by bar diagram instead of a line graph.

How to construct a Bar Diagram

1. First of all, constant (or independent variable) data (here major parts) are shown on the 'x' axis and variable data (here tonnage) on the 'y' axis. The bars are drawn vertically.

Three scales are assumed:

- (i) Scale for width of the bars: The width of all the bars must be equal.
- (ii) Scale for interval between two bars: The interval should be less than the width of the bars.
- (iii) Scale for variable data : It is to be shown on 'y' axis. It should be in round



Notes

- figures. The principle of selecting scales is the same as in case of line graph.
2. Now length of the bars are calculated as per the scale and the data.
 3. Then points are marked and bars are drawn.
 4. Labelling of the diagram is done in the same manner as is done in the line graph.

Example

Population of the major states in India is given below for 1991. Represent the graphical data using the bar diagram.

Table 5.9

<i>Sl.No.</i>	<i>State</i>	<i>Population (in Million)</i>
1.	Andhra Pradesh	66.5
2.	Assam	22.4
3.	Bihar	86.4
4.	Gujarat	41.3
5.	Haryana	16.5
6.	Himachal Pradesh	5.2
7.	Jammu & Kashmir	7.7
8.	Karnataka	45.0
9.	Kerala	29.1
10.	Madhya Pradesh	66.2
11.	Maharashtra	78.9
12.	Orissa	31.7
13.	Punjab	20.3
14.	Rajasthan	44.0
15.	Tamil Nadu	55.9
16.	Uttar Pradesh	139.1
17.	West Bengal	68.1

Source: Census of India 1991

The above data is plotted on a graph paper by bar diagram as shown in (fig. 5.3). You will note that the states are plotted on x-axis in the order these are given in the table. Equal distance from one state to another state has no meaning here. It is only a nominal distance we have taken to separate anyone state from any other



Notes

state. On the y-axis, however, the height of the bars in the case of each state is in proportion to their population and it has a scale (half centimeter is equal to 10 million population).

Some time the states can be rearranged on the basis of ascending or descending order of the population size as per the convenience.

The bar diagrams discussed above are known as vertical bar-diagrams since the bars are shown vertically. We can show these bars as horizontal bars also. In that case it will be known as horizontal bar diagram. In horizontal bar diagram vertical scale i.e. y-axis, will be nominal and horizontal scale i.e. x-axis will be a numerical scale.

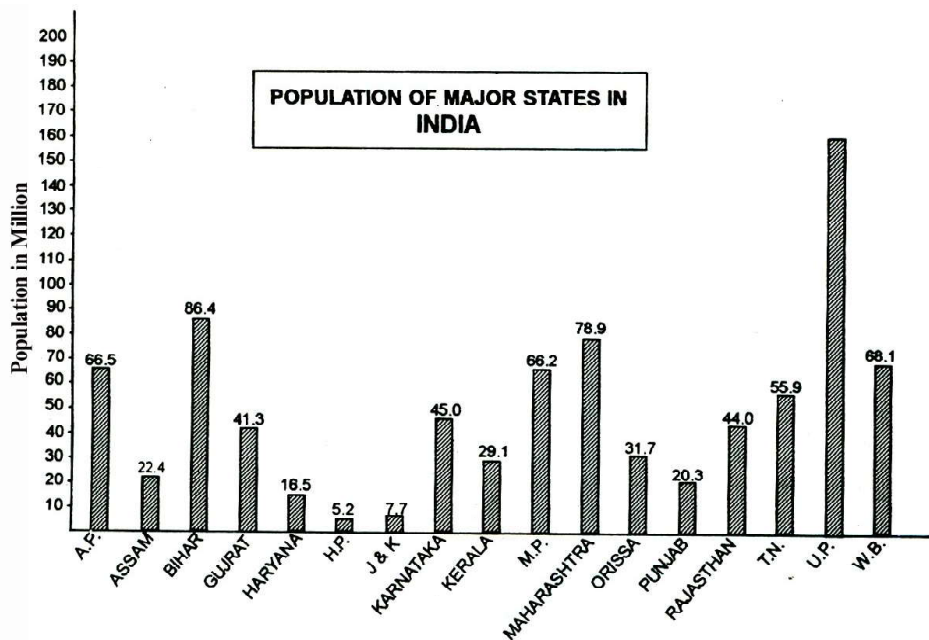


Fig. 5.3 Bar Diagram

Compound Bar Diagram

Quite often the variable being shown by the bars may consist of few different categories. These categories can also be shown on the bar itself. Such a bar diagram shows the magnitude of different values as well as share of its different categories and is known as compound bar diagram. It is also known as stacked bar diagram. In a compound bar diagram along with the magnitude bars also show the share of different categories of the variable which is shown by the bars.

Example

The population of the major states of India is given below with rural and urban



Notes

breakup. Plot the data on a compound bar diagram showing rural and urban break up of each state.

Table 5.10 Population (in million)

Sl.No.	State	Rural	Urban	Total
1.	Andhra Pradesh	48.6	17.9	66.5
2.	Assam	19.9	2.5	22.4
3.	Bihar	75.0	11.4	86.4
4.	Gujarat	27.1	14.2	41.3
5.	Haryana	12.4	4.1	16.5
6.	Himachal Pradesh	4.7	0.5	5.2
7.	Jammu & Kashmir	5.9	1.9	7.7
8.	Karnataka	31.1	13.9	45.0
9.	Kerala	21.4	7.7	29.1
10.	Madhya Pradesh	50.8	15.4	66.2
11.	Maharashtra	48.4	30.5	78.9
12.	Orissa	27.4	4.3	31.7
13.	Punjab	14.3	6.0	20.3
14.	Rajasthan	34.0	10.0	44.0
15.	Tamil Nadu	36.8	19.1	55.9
16.	Uttar Pradesh	111.5	27.6	139.1
17.	West Bengal	49.4	18.7	68.1

Source: Census of India 1991

Construction of compound bar diagram is not very much different from the ordinary bar diagram. In the final form the bars are divided into their categories using the same scale. These categories are indicated in the index, The above data is plotted by a bar diagram and is shown in fig. 5.4.

In the present case only two categories are there. In some other cases there may be several categories. In such cases each bar will be subdivided into several categories. All these categories, however, will have to be shown in an index.

In some cases the differences in absolute values may not be as important as their



Notes

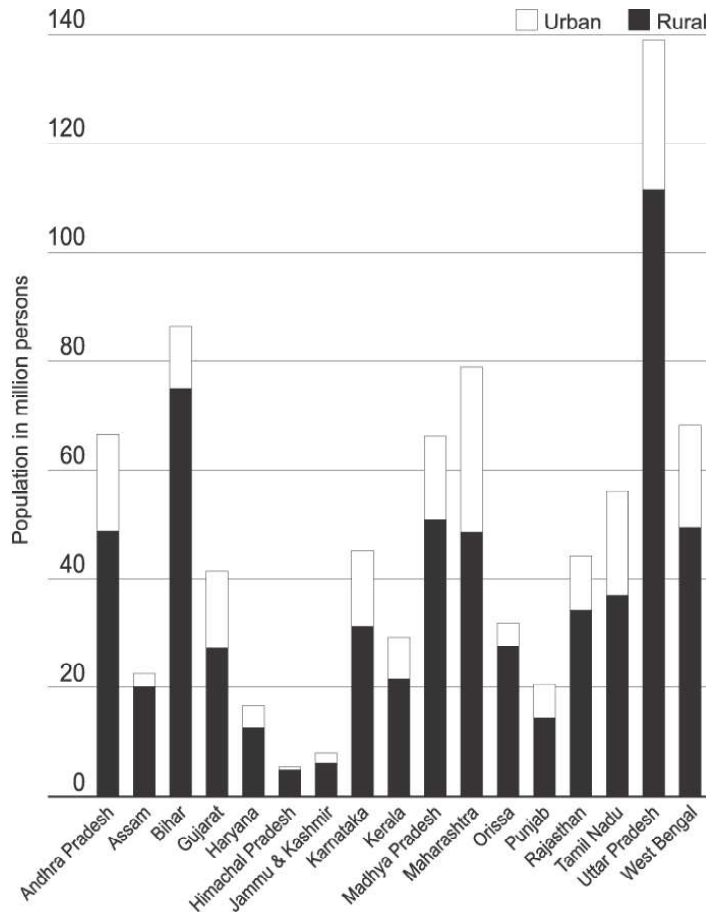


Fig. 5.4 Compound Bar Diagram

Multiple Bar diagram

Many a times it is more useful to plot sub-categories of a variable side by side. In these cases the bar of one category is placed alongwith the bars of other categories for each area. Each category is shown by a separate shade and is given in the index. The advantage of multiple bar diagram over compound bar diagram is that in this case comparison is direct. In compound bar diagram the categories are placed one over the other, whereas in multiple bar diagram these are placed side by side making comparison more direct and quicker. When there are many categories and less observations/units, multiple bar diagram is more appropriate. On the other hand, there are more observations/units and lesser categories, a compound bar diagram is always better.



Notes

Example

Electricity sold for different purposes in four states of India is given below for 1994-1995. Plot the data on multiple bar diagram and compare its use for different purposes.

Table 5.11 Electricity used (in crores kw/h) for different purposes in selected states

Sl. No.	State	Domestic Purposes	Commercial Purposes	Industrial Purposes	Other	Total
1.	Andhra Pradesh	332.0	68.3	754.9	1208.6	2363.8
2.	Bihar	73.6	42.2	637.0	219.0	971.8
3.	Maharashtra	685.3	256.5	1665.1	1481.9	4088.9
4.	Uttar Pradesh	613.3	190.5	482.7	1566.2	2552.4

Source: Statistical Abstract of India 1997

Multiple bar diagram is shown in figure 5.5 there are four categories of uses of electricity there are four bars for each state-one bar for each category. The heights of these bars are proportional to their magnitude of the electricity used (in crores kw/h.)

The multiple bar-diagram given in fig 5.5 very clearly shows that in Maharashtra, Andhra Pradesh and Bihar consumption of electricity for industrial purpose is quite high and in Uttar Pradesh it is very low. The diagram also shows that consumption of electricity for domestic and commercial purpose is strikingly low, in all the four states of India.

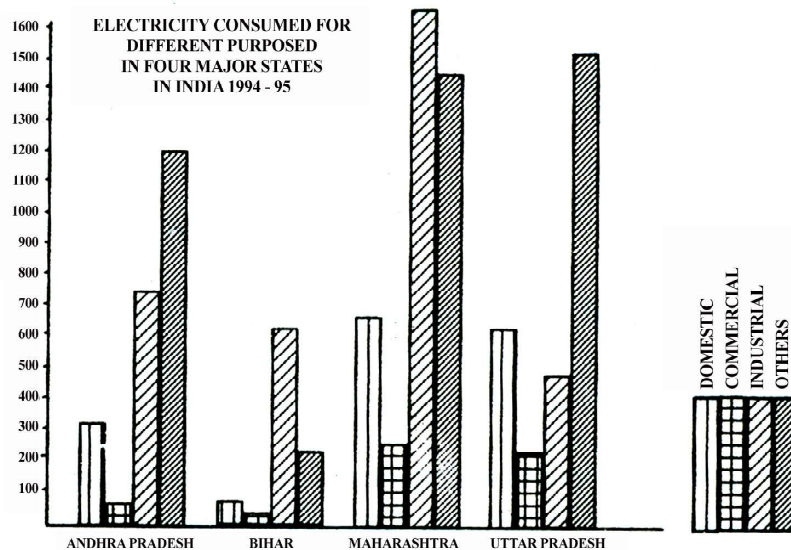


Fig. 5.5 Multiple Bar Diagram



Notes

Characteristics of a Bar Diagram

1. Bars give a visual comparison which is more effective than the quantitative comparison.
2. Multiple classifications are easily compared by a bar diagram, compound or a multiple bar diagram.
3. It can be shown on maps also.
4. It is easy to prepare and understand
5. For minor differences bar diagram or any other graph is not suitable because these are less accurate than numerical values.
6. It occupies more space.

(iii) PIE DIAGRAM

In a multiple bar diagram we compare categories of a variable for different areas. However, when number of categories increase further and number of observations are only a few, pie diagram is found to be more handy to represent these categories than the bar diagram. In a pie diagram each category is represented by different segments of a circle. The proportional share of each category is reflected in the area of the segment as well in the angle it makes at the centre.

In the construction of pie diagram one has to find out the angle of each category of the diagram. These angles are then drawn at the centre of a circle of suitable size. It will be observed that the proportional share of each category will be reflected in the area of the corresponding segment as well as in the angle.

The angle of each category is worked out by taking the ratio of the component value (C) to the total value (T) and multiplying it by 360 i.e. $(\frac{C}{T}) \times 360$. In case component values are given in percentages, each percentage is multiplied by 3.60 to convert it into corresponding angles. Sum of all such angles has to be 360° . Before making any pie diagram one should always verify this fact.

Example

Land use categories of India are given below for 1950 and 1992. Show the shift in the land use graphically with the help of a pie diagram.

Table 5.12 Land under different uses in India (in million hectare)

Year	Forrest	Non-Ag. use	Barren	Pastures	Groves	Culturable Waste	Fallow	Net Sown Area	Total
1950	40.5	9.4	38.1	6.7	19.8	22.9	28.1	118.8	284.3
1992	68.1	21.9	19.4	11.3	3.7	14.7	23.6	142.5	305.2

Source: Statistical Abstract of India 1997.



Notes

Data Interpretation and Statistical Diagrams

The land use categories given above are converted into the angles using the method given earlier.

Year	Forrest	Non-Ag. use	Barren	Pastures	Groves	Culturable Waste	Fallow	Net Sown Area	Total
1950	51.3	11.9	48.2	8.5	25.1	29.0	35.6	150.4	360
1992	80.3	25.8	22.9	13.3	4.4	17.3	27.8	168.2	360

To explain the calculation of angles let us take area under cultivable waste in 1950 which is 22.9. Its corresponding angle would be $22.9/284.3 \times 360 = 28.9975^\circ$ which after rounding upto one decimal place is 29.0. Similarly the angle of the same cultivable waste in 1992 will be $14.7/305.2 \times 360 = 17.3394^\circ$ which after rounding up to one decimal place would be 17.3 only. Note that as has been mentioned in demerits of a bar diagram, all graphical methods are less accurate than numerical methods. Rounding, therefore, upto only one or two decimal places is sufficient. Minute differences can not be depicted on the graph effectively.

The proportional composition of land use categories given in angles are shown below in fig. 5.6.

A close look of fig 5.6 will not only show proportional composition of various land use categories but also show a change that has taken place between 1950 and 1992. Pie diagram very clearly shows that forest cover has increased quite substantially between 1950-92. Also it shows a significant increase in net area sown. The diagram on the other hand shows a decline in fallow land, culturable waste, barren land and area under groves.

If we show the categories of two different areas such that one is very high and the other is low, in absolute term the size of the circle of the pie diagram can be taken in proportion to the total size. For example if we show the land utilisation of two states like Uttar Pradesh and Haryana, the size of the circle for these states may be in proportion to their area. Rest of the procedure will remain the same. One of the pie diagrams will be larger and the other will be smaller in size. Internal division of the circle, however, will show relative position of land use in two states.

Characteristics of Pie Diagram

1. It highlights the proportional composition of a phenomenon in a better way since it uses two dimensional space, unlike bar diagram which uses only heights, length for differentiating the values.
2. When there are large number of components in compound or multiple bar diagram it becomes very difficult to show them. The pie diagram is more convenience way of handling such cases.
3. It occupied lesser space as compared to bar diagram



4. It requires more mathematical calculations
5. It is effective only when proportional comparison of a few units is to be made (say two or three). In case large number of units are to be compared, a Pie diagram may not be preferred over multiple bar diagram.

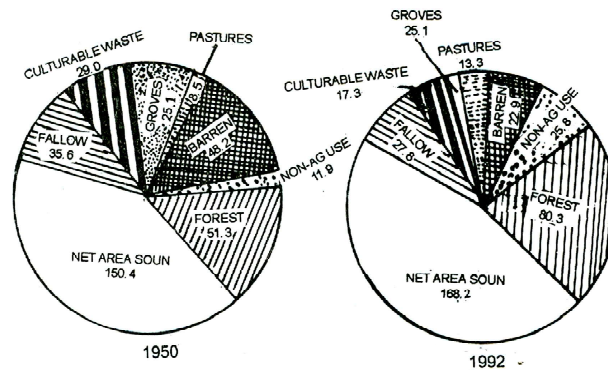


Fig. 5.6 Pie Diagram

(iv) STAR DIAGRAM

In this diagram radiating lines are drawn from a centre to represent a certain quantity or number of days etc. The length of the lines is proportional to the quantity which is to be represented. When the outer points of the lines are joined together, they give the appearance of a star. Hence, the diagram is named as a star diagram. Wind rose is a typical example of star diagrams.

Example

Construct a star diagram for the following data.

Table 5.4

Wind coming from	No. of days
North (N)	51
North East (NE)	22
East (E)	17
South East (SE)	42
South (S)	55
South West (SW)	57
West (W)	32
North West (NW)	52
Calm Days	37
Total	365

Notes



Notes

How to construct a Star Diagram

The following steps are involved in the construction of a star diagram:

- (i) There are eight directions from which wind is blowing. Hence, we draw eight lines radiating from a centre indicating all the eight directions of the wind.
- (ii) Now write eight directions on these lines as N, NE, E, SE, S, SW, W and NW.
- (iii) Assume a suitable scale for showing the flow of the wind from various directions keeping the size of the paper in view. Each line will depict the number of days the wind is blowing from each direction. Here the scale is 1 cm = 20 Days.

On the basis of the scale, the length of the lines from Centre will be calculated as under:

$$L = D \div S$$

Where L stands for length of the line.

D stands for no. of days for which wind is blowing from a direction.

S stands for scale (here 1 cm = 20 days).

Thus the length of lines of each direction will be

N	= 2.55 cm
NE	= 1.1 cm
E	= .85 cm
SE	= 2.1 cm
S	= 2.75 cm
SW	= 2.87 cm
W	= 1.6 cm
NW	= 2.6cm
Calm	= 1.85 cm

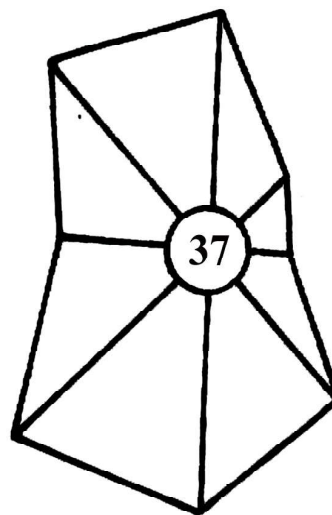


Fig. 5.7 Star Diagram

Now put points for the calculated length on each line. For the number of calm days, a circle of 1.85 cm radius will be drawn at the centre. (see fig. 5.7).

- (iv) The terminal points of each line are joined together to form the star diagram.

- (v) Number of calm days is written in the centre of the circle for which it has been drawn.

Characteristics of Star Diagram

1. Construction of star diagram is very simple. It does not involve any mathematical calculation, except the calculation of the length of lines.
2. Star diagrams are shown on climatological maps and pilot charts. They give us an understanding about the weather conditions (windy or calm) in an area of a region.

5.4 DISTRIBUTIONAL MAPS

Any phenomena or statistical data which pertains to an area is shown on the map. This map is said to be a distribution map. A variety of distributional maps may be prepared like distribution of soil, crops, population, density, literacy, rainfall, temperature, etc. For the preparation of a distribution map, following are needed—

1. An outline map of a region/area/administrative unit.
2. A relief map of the same area showing forested area/water body/marshes and contours.
3. Soil as well as climatic maps of the same area is required when crops are to be shown.
4. To show the distribution of population a map is needed on which at least urban centres are shown.

There may be a number of methods for showing the distribution but here only two methods dot and choropleth have been taken to discuss.

(A) DOT MAPS

A dot map provides a visual impression of relative density of phenomena with the help of dots of uniform size. It uses discrete data or absolute figures, which are later converted into certain number of dots.

The *dot* is a point symbol used to represent the spatial distribution of phenomena. A dot map employs either *mono-dot method* or *multiple-dot method* in representing the data. Here, we are concerned with mono dot method only.

In *mono-dot method*, the size of the dots is kept uniform throughout the region to be mapped. The dots in this case may be single coloured, when one phenomenon is to be shown, e.g., Population distribution by numbers; or multicoloured where more than one features of the same phenomena are mapped. For example, in a map showing distribution of tribes, dots of different colours indicate different tribes, but the size of dots is uniform all over, irrespective of the colour.



Notes



Notes

Example

Fig. 5.8 is a dot map showing population distribution in Haryana in 2001 (Table 5.15). In column 4 of the Table 5.15 we have worked out the number of dots to be placed in each district. The scale selected for the purpose is one dot representing 20,000 persons. The method of construction is elaborated in the following paragraphs.

Table 5.15 Population Distribution in Haryana, 2001

S.No	District	Number of Person	Number of Dots
1	2	3	4
1.	Ambala	1013660	51
2.	Kurukshetra	828120	41
3.	Karnal	1274843	64
4.	Jind	1189725	59
5.	Sonipat	1278830	64
6.	Panipat	967338	48
7.	Rohtak	940036	47
8.	Panchkula	469210	23
9.	Faridabad	2193276	110
10.	Gurgaon	1657669	83
11.	Mahendergarh	812022	41
12.	Bhiwani	1424554	71
13.	Hisar	1536417	77
14.	Sirsa	1111012	56
15.	Riwari	764727	38
16.	Kaithal	845631	47
17.	Yamuna Nagar	982369	49
18.	Fatehabad	806158	40
19.	Jhajjar	887392	44
20.	Haryana	21082989	

Source: Census of India, 2001

Construction of dot map

First of all you need a base-map of the region to be mapped and the data in



Notes

numbers. The boundaries of the individual administrative units, for which the figures are obtained are drawn in pencil or light ink. Such administrative units are called unit-areas and the value represented by each dot is the unit-value. In Example, illustrated here, 'district' is the unit-area and "20 persons per dot" is the unit-value. The preparation of dot map depends on:

- (i) Careful selection of the unit-value and size of the dots;
- (ii) Determination of uniform and proper size of dots;
- (iii) Proper placing of the dots.

It is only after the unit-value of the dot and its proper place on the map are decided, the dots of uniform size are put within the unit-areas demarcated on the map.

(i) Selection of unit-value and size of the dots

Unit value determines the number of dots to be placed in each unit area. The first step is to examine the range of quantities involved, and then to select a value to be represented by each dot. The selected unit value is always a round number (and generally a multiple of 10). Secondly, the fractions of the actual figure are never plotted.

The unit value adopted should neither be so low that it creates difficulty in inserting dots in high density areas, nor it should be so high that the area units with low density should lose their significance. The best approach is by experimenting. Dot Map in fig. 5.8 uses the unit value of 20,000 persons per dot.

(ii) Placing of Dots

The base maps showing physical and cultural features are of great help for the placement of dots on the maps. On the basis of these the positive and negative areas should be marked first. The positive areas are the parts of the region which are favourable for the distribution of the phenomena while negative areas are the relatively unfavourable parts of the region. In population distribution maps, for instance, these negative areas are known as *non-ecumene* areas, viz. the lands unsuitable for human settlements such as deserts, swamplands, flood plains etc.

In placing the dots, care must be taken not to leave the boundary areas blank. Care must also be taken so that the dotting does not inadvertently produce lines and clusters of dots that do not occur in reality. In fig. 5.8, dots have been placed after a careful study of the districtwise distribution of population in Rajasthan, and the physical and cultural features of the state.

Interpretation of a dot map

As the distributional patterns on the dot map are shown by dots, theoretically one



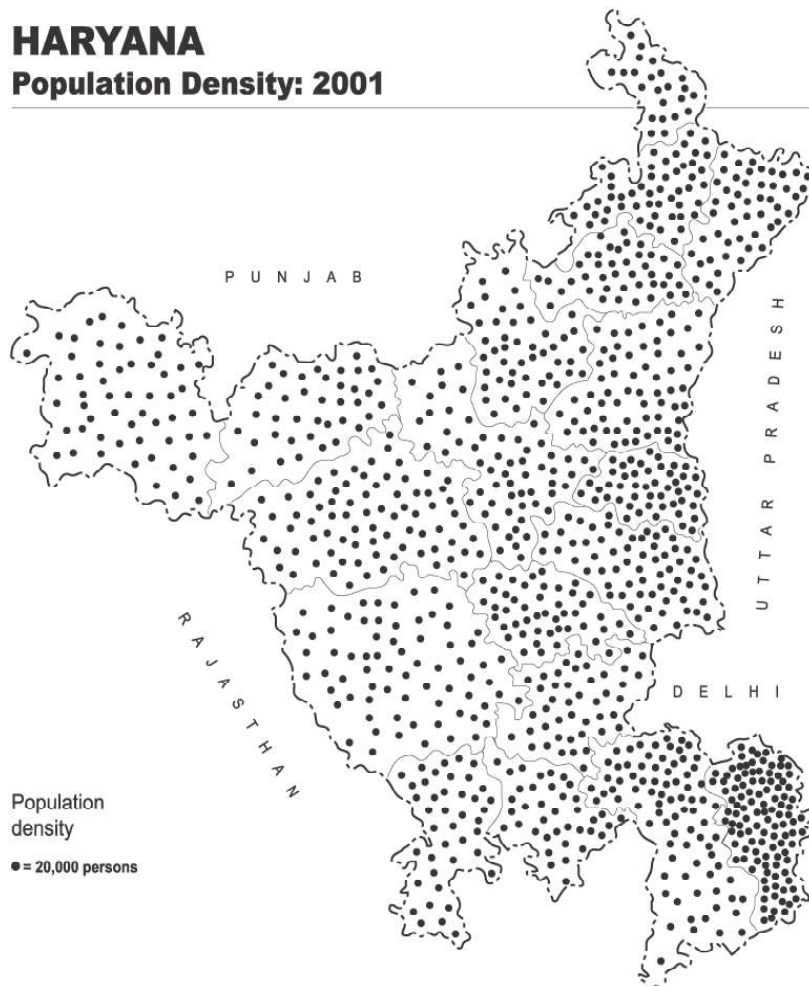
Notes

should count the number of dots in each unit area and multiply it with the unit-value assumed for one dot. The following principles should be kept in mind while interpreting a dot map:

- Divide the entire region into high, moderate and low concentration areas and describe each of them separately.
- The areas or districts not conforming to the general pattern can be discussed as exceptional cases.
- If necessary, the facts of distribution may be supplemented by the absolute figures given in the respective table.

The above principles may, now, be used to interpret the patterns of distribution of rural population in Haryana as shown in fig. 5.8. The distribution is almost even in nature. Because of gentle variation in the data, the nodes of very high concentration do not exist on the map, but the general pattern reveals the fact clearly that the south-east region of Haryana is more crowded than the north-west part. In terms of the total number of dots, the population appears to be the highest in Faridabad district. The lowest number of rural people appears to be in Panchkula.

HARYANA Population Density: 2001





Notes

Characteristics of a dotmap

1. A dot map exhibits more vivid picture of the distributional pattern.
2. It is easily commensurable.
3. A dot map may be converted into an isopleth or choropleth map, but the reverse is not possible.
4. The dot method is sometimes referred to as the “absolute method”, because of the absolute ratio between quantities represented and the number of dot employed.
5. Dot maps are comparatively easier to construct. No much computation is required in determining the number of dots required.
6. The dot may not be near the place where feature to be mapped actually exists.
7. Dot method fails to show the distribution of a phenomenon having very uneven distribution.
8. Sometimes the dots coalesce in denser areas and are rarely counted. For precise information one has to turn to the basic source of data.

(B) CHOROPLETH MAP

Choropleth is a technique for representing spatial data on map in which shadings are drawn depending upon the intensity/density of information. The raw spatial data is analysed first and different categories are made. An uniform shade/colour is assigned to one category/class. The shades are selected based on certain norms. In general darker shades are allotted to higher values/density/intensity and that of lighter shades to lower values/density/intensity.

Example

Figure 5.9 is a map showing the density of population in Haryana (Table 5.16). Column III and IV show the area of districts and their population, respectively.

The density has been computed by deviding the population of a district by its respective area. Thus it is an average number of people residing in text particular district in our square km of area. The method of construction is elaborated in the following paragraphs.

Table 5.16 Population Density in Haryana, 2001

S.No. I	District II	Area (Sq.Km.) III	Population IV	Density V
1.	Ambala	1569	1013660	646
2.	Kurukshetra	1217	828120	680
3.	Karnal	2471	1274843	516



Notes

4.	Jind	2736	1189725	435
5.	Sonipat	2260	1278830	566
6.	Panipat	1250	967338	774
7.	Rohtak	1668	940036	564
8.	Panchkula	816	469210	575
9.	Faridabad	2105	2193276	1042
10.	Gurgaon	2700	1657669	614
11.	Mahendergarh	1683	812022	483
12.	Bhiwani	5140	1424554	277
13.	Hisar	3788	1536417	406
14.	Sirsa	4276	1111012	260
15.	Riwari	1559	764727	491
16.	Kaithal	2799	945631	338
17.	Yamuna Nagar	1756	982369	559
18.	Fatehabad	2491	806158	324
19.	Jhajjar	1868	887392	475
20.	Haryana	44152	21082989	478

Source : Census of India 2001

Construction of Choropleth Map

Choropleth map is the one where real information is shown, hence a map with its subdivision boundary is needed. The sub-division of map (block/district or even state) as administrative unit becomes the unit area for the construction of choropleth map.

As we know that the people live in clusters like villages, towns or cities. The distribution of population on the space surface is not uniform or continuous. The counting of heads residing in clusters are taken on the total population. If we divide the total population of a block/district by the area of a block/district it gives the density of population. A single value of an administrative boundary is an average density considered to be uniformly distributed for mapping purpose. Hence entire district is shown by that particular shade.

After getting the density/productivity/literacy they are grouped into different classes. All the administrative units falling into that class are assigned the same shading. Therefore, wherever there is greater value darker shade is used and the lighter shade is used for smaller value. In between these two ranges shades are kept in a particular order of values.



Notes

HARYANA

Population Density: 2001

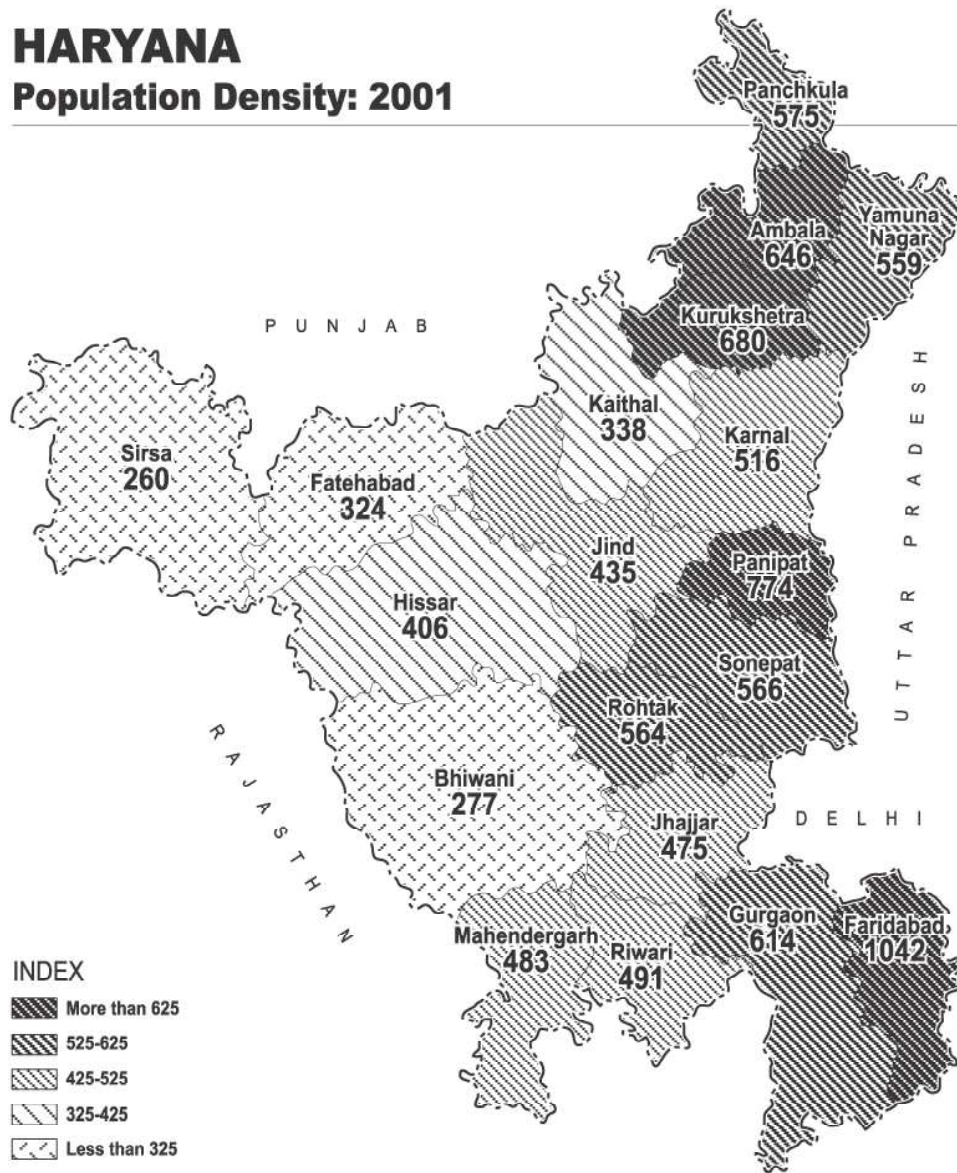


Fig. 5.9 Choropleth Map

Interpretation of choropleth map

Before making any statement about the interpretation of the choropleth map, one should always keep in mind the general pattern of shading. Pattern may be seen in terms of increasing or decreasing trend on the surface. Sometimes, a few isolated pockets of high concentration of density may be attributed due to some other favourable conditions. Hence, it may be explain in that perspective.

Just by giving a glance on the figure 5.9 it is quite clear that the higher concentration of population density is in the eastern part of Haryana. As one goes from east to

**Notes**

west the density decreases and to the north, west and south-west, density is the lowest. Excessively high density (1042 person per Km²) is found in Faridabad, located on the south-eastern most corner. The density map shows five categories of density concentrations; very low, low, moderate, high and very high. The minimum density (260 persons/Km²) is recorded in Bhiwani district in south western part of Haryana. The difference between the highest and lowest is the more than 750 persons/Km². The high density districts are adjoining to the very high density districts of Haryana. The list of districts falling under different density categories are as follow –

Very High Density – Faridabad, Panipat, Kurukshatra and Ambala;

High Density – Gurgaon, Panchkula, Sonipat, Rohtak and Yamuna Nagar;

Moderate Density – Karnal, Riwari, Mahendergarh, Jhajjar and Jind;

Low Density – Hissar and Kaithel;

Very low Density – Fatehabad, Bhiwani and Sirsa.

Characteristics of a Choropleth Map

1. Choropleth map shows different shadings. The darkness of the shade exhibits the changing scale of the value in an ascending or descending order.
2. The shade of a class / category exhibits the same values, though, the range may vary a lot.
3. The entire administrative unit considered to be uniform in terms of distribution but in reality it may not be true.
4. The administrative boundary represents a sharp divide between adjacent unit.
5. The darkest and the lowest shades may be adjacent to each other.
6. Many a time it may not be able to show the varying trends on the earth surface.
7. Pockets/Patches may appear on the map distinctively.
8. Varying trend can be compared by preparing two sets of map of two time period.

EXERCISES FOR PRACTICAL RECORD BOOK

1. Draw a line graph showing growth of population in India from 1901-91.

Years	1901	1911	1921	1931	1941	1951	1961	1971	1981	1991
Population (in million)	23.8	25.2	25.1	27.9	31.9	36.1	43.9	54.8	68.6	84.4



2. Construct a bar diagram showing Annual Rainfall in Thiruvananthapuram.

Months	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Rainfall (in mms)	22.9	20.8	38.6	105.7	207.8	356.4	223.0	145.5	137.9	273.3	205.5	74.5

3. Prepare a star diagram to represent the following data:

Direction	No. of days the wind is blowing
N	45
NE	110
E	25
SE	27
S	23
SW	15
W	90
NW	20
Calm days	10
Total	365

Notes