Chapter- 1 Data – its source and compilation

We have seen and used various forms of data. For example, at the end of almost every news bulletin on Television, the temperatures recorded on that day in major cities are displayed. Similarly, the books on the Geography of India and the world show data relating to the growth and distribution of population, and the production, distribution and trade of various crops, minerals and industrial products in tabular form.

Meanings of Datum & Data:
Datum is a single measurement.
The numerical information is called data. The data are defined as numbers that represent measurements from the real world. For example - 20 centimeters of continuous rain in Barmer or information such as New Delhi – Mumbai distance is 1385 kilometers by train.

Need of Data
1. The distribution and growth of phenomena is explained through the data in tabular form.
2. Data explains the interrelationship between many phenomena over the surface of the earth.
3. Many variables that influence the relationships can be explained best by data.
4. Statistical analysis of those variables is explained by data.
5. For example, to study the growth of a city we must have data on total population, density, number of migrants, occupation of people, their salaries, industries, means of transportation and communication, etc.
6. Thus, data plays an important role in geographical analysis.

Presentation of the Data
To avoid statistical fallacy (getting wrong conclusion), it is very important to collect the data to know the facts and figures, but equally important is the presentation of data. The data has to be analysed using statistical methods and presentation correctly.

Sources of Data
The data are collected through the two ways.
1. **Primary Sources**: - The data, which are collected for the first time by an individual or the group of individuals, institution/organisations, are called Primary sources of the data.
2. **Secondary Sources**: - The data collected from any published or unpublished sources are called Secondary sources.
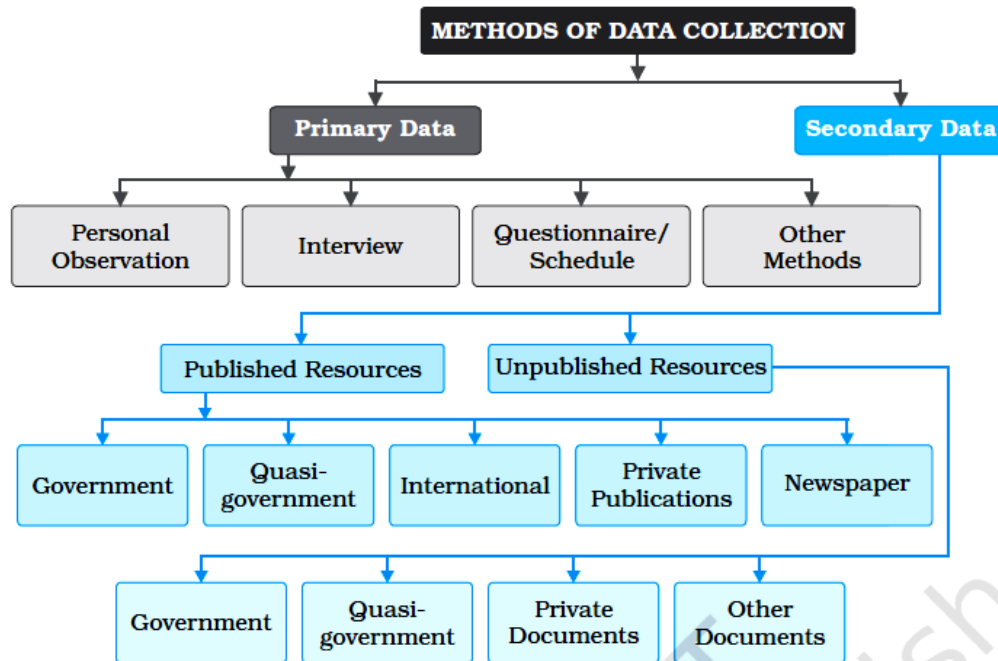
Fig. 1.1 : Methods of Data Collection

Sources of Primary Data

1. Personal Observations

It is collection of information by an individual through direct observations in the field. An example of direct observation is a field survey through it information about the relief features, drainage patterns, types of soil and natural vegetation, as well as population structure, sex ratio, literacy, means of transport and communication, urban and rural settlements, etc. is collected. The observer must have a scientific attitude for unbiased evaluation and recording.

2. Interview

In this method, the researcher gets direct information from the respondent through dialogues and conversations.

Precautions while conduction an interview:

The interviewer must take the following precautions while conducting an interview with people of the area:

(i) A precise list of items about which information is to be gathered from the persons interviewed be prepared.

(ii) The person(s) involved in conducting the interview should be clear about the objective of the survey.

(iii) The respondents should be taken into confidence before asking any sensitive question and he/she be assured that the secrecy will be maintained.

(iv) A congenial atmosphere should be created so that the respondent may explain the facts without any hesitation.

(v) The language of the questions should be simple and polite so that the respondents feel motivated and readily agree to give information asked for.

(vi) Avoid asking any such question that may hurt the self – respect or the religious feelings of the respondent.

(vii) At the end of interview, ask the respondent what additional information he/she may provide, other than what has already been provided by him/her.

(viii) Pay your thanks and gratefulness for sparing his/her valuable time for you.

3. Questionnaire/Schedule

Questionnaire: In this method, simple questions and their possible answers are written on a plain paper and the respondents have to tick-mark the possible answers from the given choices. Or, sets of structured questions are written and sufficient space is provided in the questionnaire where the respondents write their opinion.

This method is useful in carrying out the survey of a larger area. Even questionnaire can be mailed to faraway places. The limitation of the method is that only the literate and educated people can be surveyed to provide the required information.

Schedule: It is similar to the questionnaire, but the only difference between the questionnaire and the schedule is that the respondent himself/ herself fills up the questionnaires, whereas a properly trained enumerator himself fills up schedules by asking question addressed to the respondents. The main advantage of schedule over the questionnaire is that the information from both literate and illiterate respondents can be collected.

4. Other Methods

The data about the properties of soil and water are collected directly in the field by measuring their characteristics using soil kit and water quality kit. Similarly, field scientist collect data about the health of the crops and vegetation using transducers

Secondary Source of Data

Secondary sources of data consist of published and unpublished records which include government publications, documents and reports.

A. Published Sources

1. Government Publications

The publications of the various ministries and the departments of the Government of India, state governments and the District Bulletins are one of the most important sources of secondary information. These include the Census of India published by the Office of the Registrar General of India, reports of the National Sample Survey, Weather Reports of Indian Meteorological Department, and Statistical Abstracts published by state governments, and the periodical reports published by different Commissions. (Paste cover pictures of any two of these.)

2. Semi/Quasi-government Publications

The publications and reports of Urban Development Authorities and Municipal Corporations of various cities and towns, Zila Parishads (District Councils), etc. fall under this category. (Paste cover pictures of any two of these.)

3. International Publications

The international publications comprise yearbooks, reports and monographs published by different agencies of the United Nations such as United Nations Educational, Scientific and Cultural Organisation (UNESCO), United Nations Development Programme (UNDP), World Health Organisation (WHO), Food and Agricultural Organisation (FAO), etc. Some of the important publications of the United Nations that are periodically published are Demographic Year Book, Statistical Year Book and the Human Development Report. (Paste pictures of any two of these.)

4. Private Publications

The yearbooks, surveys, research reports and monographs published by private organisations fall under this category.

5. Newspapers and Magazines

The daily newspapers and the weekly, fortnightly and monthly magazines serve as easily accessible source of secondary data. (Paste pictures of any two of these.)

6. Electronic Media

The electronic media specially internet has emerged as a major source of secondary data in recent times.

B. Unpublished Sources

1. Government Documents

The unpublished reports, monographs and documents are yet another source of secondary data. These documents are prepared and maintained as unpublished record at different levels of governance. For example, the village level revenue records maintained by the patwaris of respective villages serve as an important source of village level information.

2. Quasi-government Records

The periodical reports and the development plans prepared and maintained by different Municipal Corporations, District Councils, and Civil Services departments are included in Quasi – government records.

3. Private Documents

These include unpublished reports and records of companies, trade unions, different political and apolitical organisations and resident welfare associations.

Tabulation and Classification of Data

The data collected from primary or secondary sources is in the form of raw data. To draw meaningful conclusions and to make them usable the raw data is tabulated and classified.

One of the simplest way is to summarise and present the data is the **Statistical Table**. It is a systematic arrangement of data in columns and rows. The purpose of table is to simplify the presentation and to facilitate comparisons. This table enables the reader to locate the desired information quickly. Thus, the tables make it possible for the analyst to present a huge mass of data in an orderly manner within a minimum of space.

Data Compilation and Presentation

There are three ways of data collection, tabulation and presentation in a tabular form.

1. Absolute Data

When data are presented in their original form as integers, they are called absolute data or raw data. For example, the total population of a country or a state, the total production of a crop or a manufacturing industry, etc.

Table 10.2: Mega Cities of the World

Rank	City, Country	Population in 2016 (thousands)
1	Tokyo, Japan	38140
2	Delhi, India	26454
3	Shanghai, China	24484
4	Mumbai (Bombay), India	21357
5	São Paulo, Brazil	21297
6	Beijing, China	21240
7	Ciudad de Mèxico (Mexico City), Mexico	21157
8	Kinki M.M.A. (Osaka), Japan	20337
9	Al-Qahirah (Cairo), Egypt	19128
10	New York-Newark, USA	18604

(example)

2. Percentages and Ratio:

Some time data are tabulated in a ratio or percentage form that are computed from a common parameter, such as literacy rate or growth rate of population, percentage of agricultural products or industrial products, etc.

Table 10.1: Percentage of World's Population Living in Urban Areas

Year	Percentage
1800	3
1850	6
1900	14
1950	30
1982	37
2001	48
2017	54

(example)

3. Index Number

An index number is a statistical measure designed to show changes and comparison in variable or a group of related variables. These changes are shown with respect to time, geographic location or other characteristics. It

also compares economic conditions of different locations, industries, cities or countries.

For example: Index number is widely used in economics and business to see changes in price and quantity.

The calculation of index number is done through the simple aggregate method.

(example)

Table 1.3 : Production of Iron Ore in India

	Production (in million tonnes)	Calculation	Index Number
1970-71	32.5	$\frac{32.5}{32.5} \times 100$	100
1980-81	42.2	$\frac{42.2}{32.5} \times 100$	130
1990-91	53.7	$\frac{53.7}{32.5} \times 100$	165
2000-01	67.4	$\frac{67.4}{32.5} \times 100$	207

Source – India: Economic Year Book, 2005

How data is processed?

The processing of raw data requires their tabulation and classification in selected classes. There are some steps of processing of data:

1st Step: Grouping of Data

The raw data is grouped by determining of the number of classes intervals. The selection of the class interval and the number of classes, however, depends upon the range of raw data. Eg. 0-10, 10-20, 20-30,…etc.

2nd Step: Process of Classification

Once the number of groups and the class interval of each group are determined, the raw data are classified by a method popularly known as Four and Cross Method or tally marks. It is known as Frequency Distribution. Frequencies are classified as Simple and Cumulative frequencies. For example:

Tally Marks	Number of Individual
////	4
7H7L	5
7H7L	5
7H7L //	7
7H7L /	6
7H7L ///	10

a) **Simple Frequencies**: It is expressed by 'f' and represent the number of individuals falling in each group. The sum of all the frequencies represents the total number of individual observations in the given series. In statistics, it is expressed by the symbol N that is equal to Σ f.

b) **Cumulative Frequencies**: It is expressed by 'Cf' and can be obtained by adding successive simple frequencies in each group with the previous sum.

Table 1.6 : Frequency Distribution

Group	f	Cf
00-10	4	4
10-20	5	9
20-30	5	14
30-40	7	21
40-50	6	27
50-60	10	37

For example:

Each simple frequency is associated with its group or class. The exclusive or inclusive methods are used for forming the groups or classes.

Exclusive Method: The upper limit of one group is the same as the lower limit of the next group. That is why the method is known as exclusive method, i.e. a group is excluded of its upper limits.

Inclusive Method: In this method, a value equal to the upper limit of a group is included in the same group. Therefore, it is known as inclusive method. Classes are
mentioned in a different form in this method. Normally, the upper limit of a group differs by 1 with the lower limits of the next group.

Table 1.7 : Frequency Distribution

Group	f	Cf
0 – 9	4	4
10 – 19	5	9
20 – 29	5	14
30 – 39	7	21
40 – 49	6	27
50 – 59	10	37

For example,

In this method both upper and lower limit are included to find the frequency distribution.

Frequency Polygon
A graph of frequency distribution is known as the frequency polygon. It helps in comparing the two or more than two frequency distributions. The two frequencies are shown using a bar diagram and a line graph respectively.

Ogive
When the frequencies are added they are called cumulative frequencies and are listed in a table called cumulative frequency table. The curve obtained by plotting cumulative frequencies is called an Ogive.
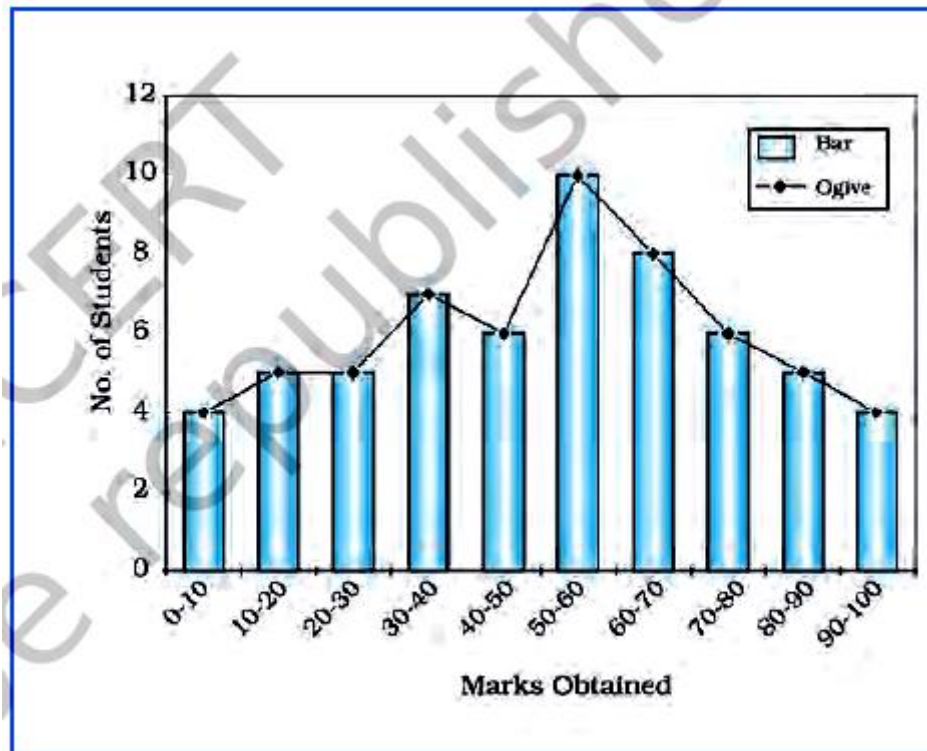

Fig. 1.5 : Frequency Distribution Polygon