Chapter 1

Data Representation techniques
Chapter 1

Data Representation techniques

This chapter will provide an introductory knowledge about the purpose and importance of various methods of representation. The theory given in this chapter seems purely related and dependant on mathematical and statistical definitions, but it is quite relevant to the real world, as each and every field of study must need collection of data and when this collection grows too large then an easy and effective means of representation is required. Some basic concepts of data representation will be discussed in this chapter. This chapter also makes the mechanism of several representation models understandable.

1.1 Introduction

Representation of data is the base for any field of study [3]. Whenever collection of data is started and the range of data increases rapidly, an efficient and convenient technique for representing data is needed. In any organization top management persons and other concerned top level authority members do not have enough time to go through whole reports regarding the progress of their firm or organization, but any small point of data should not remain hidden from their eyes. Otherwise this will affect the decisions taken by the authority people in a negative way. Therefore it is required for presenting the data in such a manner that enable reader to interpret the important data with minimum efforts and time. Several representation techniques have been developed in this concern [17]. Data presentation and data representation are terms having similar meaning and importance. Techniques for data presentation are broadly classified in two ways:

1. Non graphical techniques: Tabular Form, Case Form

2. Graphical techniques: Pie Chart, Bar Chart, Line Graphs, Geometrical Diagrams

1.1.1 Non-Graphical Techniques

There are two types of non-graphical techniques:

(i) Tabular form

(ii) case form

(a) Tabular Form:
This is better known as numerical data tables. Tabular form is the most commonly used technique for data presentation. This technique provides correlation or measurement of two
values/variables at a time. One variable is represented in columns and other in rows. Combination of row and column is known as cell, data is stored in cells and can be retrieved with the row and column notations [24, 22]. Simple example of tabular form is the representation of marks of students of a class in different subjects as shown in Fig 1.1.

![Fig 1.1: Example of tabular representation [27]](image)

(b) Case Form
This technique is rarely used. Data is presented in the form of paragraphs and follows a rigid protocol to examine limited number of variables. The protocol can be identified according the following cases:

1. Determine all the possible strategies for the present scenario.
2. Determine the optimum suitable strategy.
3. Evaluating the selected strategy in a particular case.

1.1.2 Graphical Techniques
Non graphical techniques exhibits some problems and are not so helpful when our objective is to make some decisions after analyzing set of data. So there is an alternative method of representing data which used different kind of pictorial notations. In this section we are introducing several graphical representation techniques with their importance. Graphical Techniques of representation may vary according to the need. In the field of mathematics several graphical representation techniques have been introduced by mathematicians. Similarly field of computer science also have several types of graphical representation like flow chart (a pictorial notation for the solution of any problem), E-R diagram, DFD, state diagram etc. Some of the most widely used graphical representations are shown in Fig 1.2. [25, 36]
(a) Flow chart

(b) E-R Diagram
1.2 Purpose of Data Representation Techniques

The purpose of data representation [25, 27, 29] is for recording of the information and transmitting information across two or more ends. Computer science is a science stream that involves several experiments and their planning. This discipline also collects and organizes data, evaluate the results and present its findings to public or concerned person. Therefore as a researcher (related to science stream) one need to develop the skills for
researching information, designing experiments then analyzing and presenting the data produced. The produced data can be presented in various patterns like text, table or graph. Each pattern has some merits and demerits. It is commonly said that “a graph is worth a thousand words” so we can understand the importance of graphs in scientific community. Among all methods the graphical method is the most effective way to describe, explore and summarize a set of data. In modern society graphs are extensively used to represent trend, profit & loss, quantities, sales figure, growth & decline, comparisons, ratios etc. Statistical methods[19] in combination with graphical representation form a powerful tool to analyze and communicate information (processed data). Purpose of presenting data is to get the message across and to allow others to judge your work. Methods of data representation include figures and tables in reports and paper, figures and tables in oral presentation, figures and tables in posters and websites etc. The three methods seems to be same but these are not necessarily the same. Best practice for data representation has resulted in conventions for all of these. There is no strict rule which representation one should use but generally information in figures is easier to assimilate than in tables. Tables suit lots of data that need not be assimilated but only needs to be presented. Tables suit non-numerical data as well as numerical data. Care should be taken while drawing figures for representing information. Different kind of figures are used for different kind of set of data. So it is necessary to identify that what type of data we are going to represent, before applying a particular type of pattern. Ambiguity, clarity, format, accuracy and guidelines for representing data are the general points of consideration one should keep in mind while thinking about representation of data. All data must be presented in a way that says what it means. Data in figures and legends must match text. Normally raw data are not presented in the reports, rather analyzed data is used. Obviously the analyzed data must be in a clear format. Author should use appropriate fonts, figures, legends, symbol size etc. There must be some guidelines defined for representation of data and there should not be any problem with graphics and statistics program.

1.3 Fundamental of Representation Techniques

1.3.1 Meaning of Data
Everyone might be reading a newspaper regularly. Almost every newspaper gives the weather recorded in the city on the previous day. It also indicates the rainfall recorded, and the time of sunrise and sunset. In any institution attendance of its members is being taken regularly and recorded in a register. For an athlete, the coach advises recording of the performance of the athlete at regular intervals. If we record the weather, or rainfall, or the time of sunrise and sunset, or attendance of employees, or the performance of the athlete, over a period of time, what we are recording is known as data.[45]

So, data refers to the set of observations, values, elements or objects under consideration. Data also refers to the known facts or things used as basis for inference or reckoning facts, information, material to be processed or stored. Data is unprocessed set of facts and figures, after processing data turned into information. We can say information is processed data. For example

IS RED THE ROSE

The above line does not have any meaning but it is a collection of English words hence it is data. Now look at the following

THE ROSE IS RED.

This is a sentence which comes into existence while the rules of English grammar are applied on the gathered data above. Data has no meaning until it is processed by any means. Processed data is information which has some sense or meaning.

1.3.2 Nature of Data

Data is important not only in the scientific field but persons of each and every area has to gather data hence produce some information. The nature of data varies according to its field. Following are various forms of data:

**Qualitative and Quantitative Data:**

Qualitative data are also known as categorical data because such items are identified according to its category(or some property which is the base criteria for the collection of data). For example candidates of any competitive exams can be divided into four categories SC,ST,OBC and GEN. We can gather the related data depending upon the quality of data .Quantitative data is also known as numerical data like categorization of students according to their percentage.
Continuous and Discrete Data:
If data can take on any value between two specified values it is called continuous otherwise it is called discrete. A discrete variable is one with a well defined finite set of possible values, called states. A continuous variable is one which can take on a value between any other two values, such as: indoor temperature, time spent waiting, water consumed, color wavelength, and direction of travel. A discrete variable corresponds to a digital quantity, while a continuous variable corresponds to an analog quantity.

Primary and Secondary Data:
The data collected by or on behalf of the person or people who are going to make use of the data refers to primary data. For example, the attendance of children, the result of examinations conducted by any institution are primary data. Actually, when an individual personally collects data or information pertaining to an event, a definite plan or design, it refers to primary data. Care should be taken while collecting primary data as this is the back bone for all data bases. Sometimes people may use the data already collected by someone else, such as the school attendance of children, or performance of students in various subjects, etc, for determination of level of an academic institution, such data are secondary data. The data used by a person or people other than the people by whom or for whom the data were collected refers to secondary data. For many reasons we may have to use secondary data, which should be used carefully, since the data could have been collected with a purpose different from that of the investigator and may lose some detail or may not be fully relevant. For using secondary data, it is always useful to know the method of data collection and processing, the level of accuracy, interpretation of data, level of consistency of data [17, 19].

1.4 Various Classes of Graphical Representation Techniques
This section elaborates various categories of data representation [23]. The type or class of representation for data may vary according to the use and context of data. Sometimes it is helpful to use case form of representation (which is purely theory dependant) on the other hand if computation of data is needed than tabular form is useful. While some analysis should be performed than these non graphical form of representations are not of much use. In such analysis scenario graphical form of representation is very helpful and use of graphical representation
definitely speed up the analysis process. Some forms of graphical representation and visualization are as follows:

- Graphs
- Charts
- Diagrams
- Time Series
- Data Maps & GIS
- Narrative Graphs
- Animation
- Virtual Reality

### 1.4.1 Graphical Representation of Data

The data which has been represented in the tabular form, can be displayed in pictorial form by using a graph. A graphical presentation is the easiest way to depict a given set of data. A graphical representation is a visual display of data and statistical results. It is often more effective than presenting data in tabular form. There are different types of graphical representation and which is used depends on the nature of the data and the type of statistical results. Graphical representation is the visual display of data using plots and charts. It is used in many academic and professional disciplines but most widely so in the fields of mathematics, medicine and the sciences. Graphical representation helps to quantify, sort and present data in a method that is understandable to a large variety of audiences. Several types of mediums are used for expressing graphics, including plots, charts and diagrams. Visualization techniques are ways of creating and manipulating graphical representations of data. We use these representations in order to gain better insight and understanding of the problem we are studying - pictures can convey an overall message much better than a list of numbers. In literature we found that words diagram, chart and graph are commonly being used interchangeably. But meaning of these words is as follows: [7, 42, 43 ]

**Diagram:** A diagram can be defined as a figure generally consisting of lines, made to accompany and geometrical theorem, mathematical demonstration etc. A drawing, sketch or plan that outlines and explains the parts of something, is also a type of diagram. For example a diagram of an engine. Pictorial representation of a quantity or of a relationship is termed as
diagram in simple words. An entity that can be observed and viewed easily is called diagram. Diagrams are constructed free handed and no strict rules or regulations are applied on the dimensional measurements of the object. For example a kid may draw a diagram of an apple on a complete A-4 sheet with mixture of colors, while an artist may draw the same object with some other measurements and color textures.

**Chart:** Chart is a formal type of diagram and there are many constraints for constructing a chart. Simply charts are used as a map designed to aid navigation by sea or air. Such charts are map showing special conditions or facts e.g. weather report. A sheet exhibiting information in tabulated or methodical form is also known as chart. Chart is a graphical representation of data as by lines, curves, bars etc. of a dependable variable e.g. temperature, price etc.

**Graph:** Another formally defined class of diagram is graph. Graph is simply a diagram in mathematical or scientific area of study. A drawing representing the relationship between certain set of numbers or quantities by means of a series of dots, lines, bars etc. plotted with reference to a set of axis is called graph. Graph can also be defined as a drawing depicting a relationship between two or more variables by means of a curve or surface containing only those points whose coordinates satisfy the relation. In term of computer science graph is a network of lines connecting some points [5].

In mathematics, line and dot plots are especially valuable for displaying information. Line plots, which are number lines with the letter "x" placed above numbers to show their frequency, are used to represent numerical data. These plots generally appear as straight horizontal lines: they are used to convey various kinds of data but are most useful when representing a single group of data with less than 50 values. Steam and leaf plots are other types of graphs used to present mathematical data: they are similar to line plots but display data in vertical lines rather than along horizontal axes. Stem and leaf plots are ideal for visually displaying means and averages as well as highlighting outliers and anomalies. Histograms are other common methods of graphical representation: these charts represent frequency distribution using horizontal and vertical lines.

There are variety of ways for graphical representation. Few commonly used graphical representations of data are listed below:
Types of graphical representation

(i) Histogram

A Histogram is a vertical bar chart that depicts the distribution of a set of data. Unlike Run Charts or Control Charts, a Histogram does not reflect process performance over time. It's helpful to think of a Histogram as being like a snapshot, while a Run Chart or Control Chart is more like a movie. Histogram is the most common form of graphical presentation of data. When we are unsure what to do with a large set of measurements presented in a table, we can use a Histogram to organize and display the data in a more user-friendly format. A Histogram will make it easy to see where the majority of values falls in a measurement scale, and how much variation there is. Histogram is helpful when we want to do the following:

Summarize large data sets graphically. We can see that a set of data presented in a table is not easy to use. We can make it much easier to understand by summarizing it on a tally sheet and organizing it into a Histogram. Compare process results with specification limits. If we add the process specification limits to our Histogram, we can determine quickly whether the current process was able to produce "good" products. Specification limits may take the form of length, weight, density, quantity of materials to be delivered, or whatever is important for the product of a given process. Communicate information graphically. The team members can easily see the values which occur most frequently. Use a tool to assist in decision making.

Method of plotting histogram:

For plotting a histogram, one has to take a graph paper. The values of the variable are taken on the horizontal axis scale known as X-axis and the frequencies are taken on the vertical axis scale.
known as Y-axis. For each class interval a rectangle is drawn with the base equal to the length of the class interval and height according to the frequency of the C.I. When C.I. are of equal length, which would generally be the case in the type of data you are likely to handle in school situations, the heights of rectangles must be proportional to the frequencies of the Class Intervals. When the C.I. are not of equal length, the areas of rectangles must be proportional to the frequencies indicated (most likely you will not face this type of situation).

![Histogram Example](fig1.3.png)

**Fig 1.3: Example of histogram [19]**

**(ii) Bar Diagram or Bar graph or Bar chart**

If the variable is discrete, then a histogram cannot be constructed as the classes are not comparable in terms of magnitude. However, a simple graphical presentation, quite similar to histogram, known as bar graph, may be constructed.

A bar graph is a chart that uses either horizontal or vertical bars to show comparisons among categories. One axis of the chart shows the specific categories being compared, and the other axis represents a discrete value. Some bar graphs present bars clustered in groups of more than one (grouped bar graphs), and others show the bars divided into subparts to show cumulative effect.

**How to use it:**
Determine the discrete range. Examine your data to find the bar with the largest value. This will help you determine the range of the vertical axis and the size of each increment. Then label the vertical axis.

Determine the number of bars. Examine your data to find how many bars your chart will contain. These may be single, grouped, or stacked bars. Use this number to draw and label the horizontal axis.

Determine the order of the bars. Bars may be arranged in any order. (A bar chart arranged from highest to lowest incidence is called a Pareto chart.) Normally, bars showing frequency will be arranged in chronological (time) sequence.

Draw the bars. If you are preparing a grouped bar graph, remember to present the information in the same order in each grouping. If you are preparing a stacked bar graph, present the information in the same sequence on each bar. Label and title the graph.

![Example of bar diagram](image)

**Fig 1.4: Example of bar diagram [19]**

### (iii) Frequency Polygon

Frequency polygon is simply a set of points joined up by straight lines. The frequency polygon is a graph that displays the data by using lines that connect points plotted for the frequencies at the midpoints of the classes. The frequencies are represented by the heights of the points. When to
**Use:** Frequency polygons are used to plot frequencies of data in different classes and are useful to show patterns and trends within the data.

**Drawing frequency polygon:**
Draw a horizontal and vertical axis. The frequency is usually measured against the vertical axis and the values of variables being measured are shown on the horizontal axis. Mark points to show the frequency in each class. Connect these points with straight lines to complete the frequency polygon.

---

**Fig 1.5: Example of frequency polygon [19]**

(iv) **Cumulative Frequency Curve or Ogive**
The next type of graph that can be used represents the cumulative frequencies for the classes is called the cumulative frequency graph or ogive. The ogive is a graph that represents the cumulative frequencies for the classes in a frequency distribution. The cumulative frequency is the sum of the frequencies accumulated up to the upper boundary of a class in the distribution.
An ogive is a graph that represents the cumulative frequencies for the classes in a frequency distribution. It shows how many of values of the data are below certain boundary.

**Steps for constructing an ogive:**

Draw and label the x (horizontal) and the y (vertical) axes.

Represent the cumulative frequencies on the y axis and the class boundaries on the x axis. Plot the cumulative frequency at each upper class boundary with the height being the corresponding cumulative frequency. Connect the points with segments. Connect the first point on the left with the x axis at the level of the lowest lower class boundary.

![An Ogive](image)

Fig 1.6: Example of an ogive [19]

*(v) Line graph or Stick graph*
Line graph or stick graph is most appropriate for comparing discrete variables and in this method of representation lines or sticks of length proportional to the frequency are used. Labeling is done in the same manner as in bar graph. Example is shown in Fig 1.8.

![Line graph example](image)

**Fig 1.7: Example of Line/stick graph [25]**

**(vi) Pie Chart**

A type of graph is which a circle is divided into sectors that each represent a proportion of whole. Each sector shows the relative size of each value. A pie chart displays data, information and statistics in an easy to read ‘pie slice’ format with varying slice sizes telling how much of one data element exists. Pie chart is also known as circle graph. The bigger the slice, the more of that particular data was gathered. The main use of a pie chart is to show comparisons. Various applications of pie charts can be found in business, school and at home. For business pie charts can be used to show the success or failure of certain products or services. At school, pie chart applications include showing how much time is allotted to each subject. At home pie charts can be useful to see expenditure of monthly income in different needs. Reading of pie chart is as easy as figuring out which slice of an actual pie is the biggest.
A Pie chart presents data as a simple and easy to understand picture. It can be an effective communication tool for even an uninformed audience, because it presents data visually as a fractional part of a whole. Users see a data comparison at a glance, enabling them to make an immediate analysis or to understand information quickly. This type of data visualization chart removes the need for readers to examine or measure underlying numbers themselves. One can also manipulate pieces of data in the pie circle to emphasize points which are needed. A Pie chart becomes less effective if it uses too many pieces of data. For example, a chart with four slices is easy to read while one with more than ten slices becomes complicated, especially if it contains many similar sized slices. Adding data labels and numbers may not help in this case, as they themselves may become crowded and hard to read. This may make it more difficult for readers to analyze and assimilate information quickly. Comparing data slices in a circle also has its problems, because the readers has to factor in angles and compare non-adjacent slices. Data manipulation within the chart’s design may lead readers to draw inaccurate conclusions or to make decisions based on visual impact rather than data analysis.

vii) Pictogram:
A pictorial symbol for a word or phrase. This class of representations is used to illustrate broad differences between categories (qualitative and discrete variables). In this type of representation simple pictures are used instead of bars to represent frequencies. A key is to be added to show what each picture represents. An example of pictogram is given in fig 1.9.

Monday

Tuesday

Wednesday

Thursday

Friday

Saturday

Sunday

Number of eggs prepared for breakfast in a hotel

represents 12 eggs

Fig 1.9: Example of pictogram [26]
viii) Line chart

Line chart is the most basic type of chart used in finance and it is generally created by connecting a series of past recorded data together with a line. It is a style of chart that is created by connecting a series of data points together with a line. Line charts are ideal for representing trends over time. A standard example would be how the stock value for a certain company develops over time on the stock market. However, it does not necessarily need to be time along the X-axis. Any data that behaves like a function with respect to the variable on the X-axis can be plotted. Line charts emphasize time flow and rate of change rather than the amount of change. Line graphs are used for both continuous and discrete data. Line graphs have ability to combine as many line series as needed. Line charts can specify whether a series should be an area chart.

![Monthly Average Temperature](source: WorldClimate.com)

**Fig 1.10: Example of Line chart [19]**

Line charts also allow the use of steps instead of a straight line as shown in Fig 1.11.
Stem-leaf diagrams are a fancy way of listing a fairly large group of numbers in order. These diagrams are seen as a quicker, more convenient and ultimately more useful way of presenting data than just a long list of numbers. For creating stem-leaf diagram part of the number, often the whole number part, is used as stem (placed vertically under each other), the other part of the number forms the leaves. Stem and leaf diagrams are used to represent ungrouped quantitative data. This type of diagrams are the only graphical representations that also display all the original data values.

**Constructing stem-leaf diagram:** A given set of data can be viewed using stem-leaf diagram. To prepare a stem-leaf diagram of a given data we divide each data value into two part. The left group is known as a stem and the remaining group of digits on the right is known as a leaf. Data can be displayed by horizontal rows of leaves attached to a vertical column of stems.

### Example of Stem-Leaf Diagram

<table>
<thead>
<tr>
<th>Test Scores</th>
<th>Stem Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>5</td>
</tr>
<tr>
<td>60,65,65,</td>
<td>6</td>
</tr>
<tr>
<td>70,70,75,75</td>
<td>7</td>
</tr>
<tr>
<td>80,85,85,85</td>
<td>8</td>
</tr>
<tr>
<td>90,95,95,95</td>
<td>9</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
</tr>
</tbody>
</table>
Stem-leaf plots are convenient to display group of data in a better way. But disadvantage of the stem-and-leaf plots is that data must be grouped according to place value. We can not use different groupings. In such case histograms are more suited. For comparing two sets of data a back-to-back stem and leaf plot is used in which the leaves of sets are listed on either side of the stem as shown in Fig 1.12.

![Stem-leaf plot](image)

Fig 1.12: Example of stem-leaf plot [7]

**x) Scatter diagram**

While working with statistical data it is often observed that there are connections between sets of data. For example the mass and height of persons are related: the taller the person the greater his/her mass. To find out whether or not two sets of data are connected scatter diagrams can be used. A scatter diagram is a tool for analyzing relationship between two variables. One variable is plotted on the horizontal axis and the other is plotted on the vertical axis. The pattern of their intersecting points can graphically show relationship patterns. Commonly a scatter diagram is used to prove or disprove cause-and-effect relationships. While scatter diagram shows relationships, it does not by itself prove that one variable causes other. In addition to showing possible cause and effect relationships, a scatter diagram can show that two variables are from a common cause that is unknown or that one variable can be used as a surrogate for the other.
**When to use scatter diagrams:** Scatter diagrams are used to examine theories about cause-and-effect relationships and to search for root causes of an identified problem. Varieties of scatter diagrams are shown in Fig 1.13.

![Example of scatter diagram](image)

Fig 1.14: Example of scatter diagram [7]

These above discussed graphical representations are purely related to mathematical aspect of any field. All disciplines which can be measured mathematically by their property can use any one of the above discussed representation methodology as and when needed. The analysis of data is a very important and basic step while studying any discipline. Collection of data requires some means for gathering data in an efficient and effective pattern. Listing is an elementary method for grouping data when data is limited and we switch to another approach (tabular approach) for grouping data while data range is high and its property is varying in nature. Tabular
representation captures more characteristics in compare to listing of data. Listing of data is linear in arrangement while tables have two dimensions which makes the tabular approach more capable for representing data in a wider range and variety. Still tables are less convenient for retrieving all the information at a single sight rather graphs are much easier and convenient. One can observe and analysis the data at a glance if data is represented pictorially. The most common and popular ways of presenting the data are tabular form or graphical form. We can easily observe the differences in tabular and graphical representation. Both of these methods have their own importance. Sometimes data can be better presented by table than by graph. For example to determine the price of milk using two axis formula based on fat percentage. This information can be effectively presented in the form of a table rather than a graph. On the other hand generally graphs are useful when there is a trend or comparison to shown.

This section has elaborated about the types of representation methods. All these methods are capable to represent data either in graphical or non-graphical form and each class of representation is designed to capture the relative properties of data. Representation techniques also aim to provide convenient means for data storage. Storage of graphical methods of data representation are complex enough with compare to tabular methods. Retrieval of data depend upon the complexity of data representation with respect to storage. Simply easy representation does not ensure fast retrieval so it is necessary to devise some tools for easy retrieval and handling of the representation pattern.

1.4.2 Terminology of Graph Theory

Graph theory concerns with mathematical and scientific representation of data and is helpful for investigating huge amount of data. Several theories were solved with the concepts of graph theory. This section will introduce terminology used in graph theory. Graph theory may have its beginning in 1736 when EULER considered the Königsberg bridge problem. It took 200 years before the first book on graph theory was written. Since then graph theory has developed into an extensive and popular branch of mathematics, which has been applied to many problems in mathematics, computer science, and other science streams. There are various types of graphs, each with its own definition.
Unfortunately, some people apply the term “graph” rather loosely, so one can’t be sure what type of graph they’re talking about unless we ask them. Examples of extensively used graph are computer networks, path of the people travelling from one city to another etc. Many of the most important molecules in biology and chemistry have complex structures that can be described with graphs [30]. The vertices in such a graph represent atoms, and the edges represent bonds between atoms. An important feature of such graphs may be the degree of the vertices i.e. the number of adjacent vertices. This corresponds to the number of bonds a particular atom has in the molecule. There are chemical properties of atoms of a given type that determine the number of bonds they typically form (e.g. 1 for hydrogen, 2 for oxygen, 4 for carbon, etc…) and it may be important that the graphs satisfy certain constraints regarding the degrees of the vertices. In order to represent double bonds between two atoms, it may be appropriate to extend our notion of a graph to a multi-graph which permits multiple edges between vertices (note that the Konigsberg Bridge diagram corresponds to a multi-graph).
One particularly important category of molecules are hydrocarbons (molecules made of carbon and hydrogen), and in such graphs an important feature is whether there are any cycles present, and if so, how many and of what size. A connected graph without any cycles is called a tree, and many saturated hydrocarbons have this structure (some small examples are shown below). In the field of nano-technology, complex carbon molecules play an important role, e.g. the molecule C60, also known as Buckminsterfullerene or a “bucky ball”. This molecule contains many cycles composed of 5 or 6 carbon atoms (it is a truncated icosahedron). Graph theory describes a number of important properties related to characterizing graphs which are trees, and analyzing the cycles in graphs which are not trees. Graphs are very useful tool for solving several programming strategies. The concepts of graph theory devise techniques for solving mathematical problems. As each and every field of study results in a collection of data so graph theory is important with regard to every subject, where collection of data occurred. Some basic terms related to graph theory are given in the following section.

**Graph:** A graph G is a discrete structure consisting of nodes (called vertices) and lines connecting the nodes (called edges). Two vertices are adjacent to each other if they are joint by an edge [70]. This edge joining the two vertices is said to be an edge incident with them. There are basically two types of graphs according to the type of edges used: directed and undirected graphs. (Fig 1.16)

![Fig 1. Undirected Graph](image1)

![Fig 2. Directed Graph](image2)

Fig 1.16: Directed and undirected graphs [69]
In directed graphs the connecting edges are having arrows indicating the direction of that particular edge. While no arrows are used in un-directed graphs. The example of directed graphs may include roads connecting various cities and having one way traffic, in this example cities are vertices and connecting roads are edges of the graph. Also direction of arrow represents the flow of traffic. While directions are necessary to add with the representation of data directed graphs are helpful. For example a travelling route must indicate the direction of tour with respect to the tourists, so directed graphs are required. On the other way if we want to only represent the distances of the travelling route then undirected graphs are sufficient for such type of representation.

Some other commonly used definitions with respect to graph are given below.

**Simple graph:** Undirected edges are used in simple graphs. Multiple edges and loops are not allowed.

**Multi graph:** A multi-graph also consists of undirected edges like simple graph. No loops are used in this type of graphs. Multiple undirected edges exist in a multi-graph.

**Pseudo graph:** A graph with undirected edges is called pseudo graph when it contains multiple edges and loops in it.

**Simple directed graph:** A simple directed graph consists of directed (arrow ended) edges. Like simple graphs multiple edges and loops are not allowed.

**Directed multi-graph:** It is similar to multi-graph, multiple edges are allowed but no loops. Only difference is that directed edges are used unlike multi-graph.

**Mixed graph:** A combination of all the above types of graph is known as mixed graph. Both directed and undirected edges are used in this type of graph. Multiple edges and loops may or may not exist.
(a) Types of undirected graph

(b) Directed graph

(c) Directed multi-graph
Fig 1.17: Example of graphs [56]

**Representation of graph**

Graph is a data structure [70] that consists of two main components: a finite set of vertices also called nodes and a finite set of ordered pair of the form \((u,v)\) called as edge. The pair is ordered because \((u,v)\) is not same as \((v,u)\) in case of directed graph (di-graph). The pair of form \((u,v)\) indicates that there is an edge from vertex \(u\) to vertex \(v\). The edges may contain weight or value or cost. Graphs are used to represent many real world applications: Graphs are used to represent networks which may include computer network or circuit network or network of organizations. Representation of graph is very important as it will affect the storage of graphs in computer memory. Graphs are capable to represent all types of objects precisely [62]. Multiple objects are treated as nodes or vertices and the type of relationship they exhibit may be treated as edges or connecting lines. A single object can also be represented using concepts of graph. We just need to identify its basic characteristics (can be used as nodes) and the inter-relationships (can be used as edges). The concepts of graph would solve the complicacy of representation and minimize the effort of representation.

Graphs can be represented in two ways: using adjacency matrix and using adjacency lists. The choice of the graph representation is situation specific. It totally depends on the type of operations to be performed and ease of use. Adjacency matrix and adjacency lists can be used for
both directed and undirected graphs. When a graph is sparse adjacency lists are preferred, on the other hand for graphs which are dense in nature adjacency matrix is better choice of representation. When graphs are represented by an adjacency matrix, operations with these graphs are faster due to the use of easy data structure i.e. arrays.

![Directed Graph (a) and Adjacency Matrix Representation (a)](image1)

![Undirected Graph (b) and Adjacency Matrix Representation (b)](image2)

Fig 1.18: Adjacency matrix representation of directed and undirected graph [56]

While adjacency list representation is itself complicated data structure so operations with the graphs represented by an adjacency lists become slower. But there is a limitation upon the size of matrix, so for large graphs adjacency lists approach of representation will be preferred. Advantage of adjacency matrix representation is that this approach is easier to implement and follow, but it consumes more space. Even if the graph is sparse adjacency matrix representation
will consume the similar space like a dense one. The another approach of representing graph is advantageous because it saves space and adding a node in such representation is easier. The problem of degeneracy could be minimized with the use of graph [72]. Numerical characterization of the objects may also be used for representation [71].

Fig 1.19: Adjacency lists representation for directed and undirected graph [56]

Adjacency list representation of graph is the base concept for the future scope of the study. Adjacency lists grow in size as and when required by adding new nodes in the linked list. Biological sequences can be represented with the help of graphs. Both the methods of representing graph will be applied to determine which approach will show better representation.

1.5 Comparison of several Representation techniques
To determine a particular representation method for data, one need to identify the type & nature of data, field of study and purpose of representation of data. Any of the existing non-graphical or graphical patterns of representation may be used as and when they suit the requirement. Depending upon the class of data tabular and graphical procedures for representing data are outlined in Fig 1.20 below.

**Outline: Tabular and Graphical Procedures**

![Outline Diagram]

Fig 1.20: Tabular & Graphical techniques of data representation [83, 144]

Differences between graphs and diagram: Actually it is hard to find out a clear cut difference between a graph and a diagram, however based on the experience and exposure to different scenario the following points of differences between a graph and a diagram can be summarized:

1. A graph paper is required to draw a graph while a diagram can be drawn on any type of paper. In technical terms we can say that graph is a mathematical relationship between two or more variables. This is not the case of a diagram.
2. Diagrams are attractive so they are used for the purpose of publicity and propaganda. On the other hand graphs are more useful for statisticians and research workers for the purpose of further analysis.

3. For representing frequency distributions, graphs are used extensively. For example for the time series graphs are more appropriate than.

4. Graphs are constructed over axis, while diagrams can be drawn in any way.

Comparisons between diagrammatic and tabular representation are listed below:

1. Diagrams and graphs are meant for a layman. Tables are meant of statisticians for the purpose of further analysis.

2. Diagrams give only an approximate idea. Tables contain precise figures. Exact values can be read from tables.

3. Diagrams can be more easily compared and can be interpreted by a layman. Comparisons and interpretation of tables can only be done by statisticians and it is a difficult task [144].

4. Diagrams and graphs cannot present much information. Tables can present more information.

5. Diagrams are more attractive and have a visual appeal. Tables are dull for a layman (may be attractive to a statistician).

1.6 Chapter Summary

This chapter has described why representation techniques are needed and how better representation can be achieved. Computer science is a discipline which encompasses all the fields of study, this chapter has justified the importance of inter-disciplinary area of study and explained the use of several concepts of computer stream. The basic idea of this research has been conceived from the concepts of graph theory. Several classes of representation techniques have been highlighted with their comparative description. In the next chapter biological sequences will be introduced with the description of their structure, component and characteristics.