ON EFFECTIVE USE OF STATISTICAL DIAGRAMS

Bipinkumar V

Department of Mathematics, Sree Narayana College,

Cherthala, Alappuzha, Kerala, India.

Abstract

A diagram is a symbolic representation of information according to visualization technique. Diagrams have been used since ancient times, but became more prevalent during the Enlightenment. A diagram can be used to explain statistical data and other important information, like how a particular system functions, quickly and with less of a strain on resources. This paper is an attempt to describe different ways of pictorial presentation of data in more attractive and evident manner. A well-constructed diagram or graph will convey much more information more effectively than a narrative-based process.

Keywords: Diagrams, Graphs, Charts.

Introduction

A diagram is essentially a picture that communicates information. That is, it is a visual form for presentation of statistical data, highlighting their basic facts and relationship. If we draw diagrams on the basis of the data collected they will easily be understood and appreciated by all. It is readily intelligible and save a considerable amount of time and energy. The construction of diagrams is an art, which can be acquired through practice. However, observance of some general guidelines can help in making them more attractive and effective. The selection of the type of diagram should be determined according to the nature of data which is being handled. The basic definitions used here are from [1] and [2].

Significance of Diagrams and Graphs

Diagrams and graphs are extremely useful because of the following reasons.

 1. They are attractive and impressive.

2. They make data simple and intelligible.

3. They make comparison possible

4. They save time and labour.

5. They have universal utility.

6. They give more information.

7. They have a great memorizing effect.

General rules for constructing diagrams

1. A diagram should be neatly drawn and attractive.

2. The measurements of geometrical figures used should be accurate and proportional.

3. The size of the diagrams should match the size of the paper.

4. Every diagram must have a suitable but short heading.

5. The scale should be mentioned in the diagram.

6. Diagrams should be neatly as well as accurately drawn with the help of drawing instruments.

7. Index must be given for identification to make out the meaning to the reader.

8. Footnote must be given at the bottom of the diagram.

9. Economy in cost and energy should be exercised in drawing diagram.

Statistical diagrams or charts can be mainly used to indicate the distribution of the data, the division of components of the data or the trends of the variation of values in the observations. Here some commonly used diagrams are discussed. [3] and [4] gives important points to be remembered in the use of diagrams.

Pie Charts

Pie charts are easy to make, easy to read, and very popular. They are used to represent categorical data or values of variables. They are basically circles that are divided into segments or categories which reflect the proportion of the variables in relation to the whole. Percentages are used to compare the segments, with the whole being equal to 100%.

To make a pie chart, draw a circle with a protractor. Then, convert the measures of the variables into percentages, and divide the circle accordingly. It is best to order the segments clockwise from biggest to smallest, so that the pie chart looks neat and the variable are easy to compare. It is also recommended to write percentage and category labels next to each segment, so that users are not required to refer to the legend each time they want to identify a segment. Pie charts are very popular type of graphs, but they do have disadvantages that limit their use. For this reason, scientists are not fans of pie charts. First of all, pie charts with too many segments look very messy and are difficult to understand; therefore it is best to use pie charts when there are less than five categories to be compared. Further, if the values of the categories are very close, the pie chart would be difficult to decipher because the segments would be too close in size. Variations of pie charts include the polar area diagrams and cosmographs.

Pictographs

Pictographs, also called pictograms, are diagrams that show and compare data by using picture symbols. Each of these symbols corresponds to a specific quantity and is repeated a number of times. The media often uses pictographs to compare trends. Schools use them, as well, in order to train students in mathematics and other subjects in an enjoyable way. Elementary level students often encounter pictographs in their textbooks. These types of graphs are also popularly used by charity organizations to track fund drives. The best example

of this is the picture of the thermometer displayed by these organizations. The thermometer represents the total goal amount and its red stripes symbolize the collected donations.

While pictographs are easy to understand, they can be misleading because they provide a general representation. Therefore, they are not commonly used by statisticians and scientists who work with very precise measurements. It is virtually impossible to accurately display the difference between $0.56, $0.61, $11.99 and $12.32 through picture symbols on the same pictograph, for example. Pictographs would be unreliable for this purpose. Sometimes the media may take advantage of the potential unreliability of pictographs and intentionally use them to exaggerate or downplay specific data, in order to influence public opinion on an issue.

Organizational charts

Organizational charts, also called organization charts or org charts, are diagrams that reveal the overall structure of the workforce of a company. Through an organizational chart, the formal indirect or direct relationships between the positions in a company are presented. The chart also shows how different departments are connected. Organizational charts are types of graphs that depict four types of relationships: line, lateral, staff, and functional. Line relationships exist between superiors and subordinates. Lateral relationships exist between different departments of similar rank. Staff relationships exist between a managerial assistant and a line manager. Functional relationships exist between a specialist and a line manager.

An organizational chart is usually shaped like a pyramid, with the President or Chief Officer in the top rectangle and levels of subordinates in descending rectangles according to rank listed below. Each rectangle size corresponds to the level of authority. Thus, superiors have larger rectangles than subordinates. Peers have equally sized rectangles. Solid lines between rectangles signify a direct relationship, and dashed lines symbolize an indirect relationship. Arrows represent the direction of communication flow between the components of organizational charts.

An organizational chart has its advantages. It promotes structure in an organization and defines the roles of the management. It also reveals the parts of a company that need improvement and possibly more or less employees. An organization chart also has disadvantages. It does not reveal anything about the managerial style. Moreover, organizational charts needs to be changed every time the statistics of employee leaving or joining the company changes.

Flowcharts

Flowcharts are types of graphs that display a schematic process. Contemporary flow charts are modelled after the logic behind early computer games. Businesses often use them to visually depict all the stages of a project. Therefore, individuals working on a project refer to a flow chart to see the breakdown of the process and understand the whole picture. A flowchart can effectively be used as a training tool for employees who are being introduced to a new project. It also helps in locating and correcting errors in a project. Even though flow chart use is usually linked to the field of business, a flowchart can be used for any purpose that involves a cross-functional process. It can even be used to show driving directions from one location to another.

A flowchart consists of start points, end points, inputs, outputs, and routes which are commonly represented by basic symbols that are labelled. Ovals represent start and end points. Rectangles represent the steps of a process, and diamonds represent decisions. Diamonds have two routes stemming from them; one is a true or yes route and the other is a false or no route. Circles stand for operations; arrow-shaped figures stand for transportation, while triangles represent storage, and squares stand for inspection. Arrows linking symbols signify the sequence of a process. While most flowcharts use shapes as symbols, others use graphics instead.

Cosmograph

Cosmographs are variations of pie charts. Like pie charts, cosmographs also provide information about components, in relation to a whole. However, cosmographs are not as numerical as pie charts. There are different uses for cosmographs. A cosmograph is commonly used to monitor finances. In this context, it is used to track the statistics of input and output of a business or organization. Government agencies, for instance, use cosmographs when they want to make a visual presentation about the money that entered the agency and how the amounts were used. Components of input, or income, are listed on the left side of the cosmograph, and components of output, or expenditure, are listed on the right side. The size of each component on a cosmograph corresponds to its value. Larger amounts mean physically larger components. Besides monetary value, an input-output cosmograph can deal with quantity or percentage. A cosmograph can also be used to show a comparison between geographical regions. Publications frequently feature cosmographs that use different colours to provide information about regions, in respect to a whole.

Scatter plot charts

A scatter plot chart simply plot the observations of the data in a paper to give a collection of scattered points. Even though a scatter plot gives collection of scattered points, a lot of information can be obtained from it. For example, it gives a rough idea of inter-variable relationship between the components such as how the increase in the value of one variable is affected in the value of the other variable. We can even fit a best fitting curve to the set of observations. Obtaining the regression lines and finding the correlation between the variables can also be applied to the set of observations. Word clouds are an interesting example of scatter plot charts which are used to visualize the frequency distribution of words with textual data. A word cloud clearly emphasise the words which are used several times in a particular talk. For example, people may be eager to know which word was used several times in a talk delivered by the President or in some academy award function.

Bar graphs

Bar graphs are used to present and compare data. There are two main types of bar graphs: horizontal and vertical. They are easy to understand, because they consist of rectangular bars that differ in height or length according to their value or frequency. These types of graphs serve the same purpose as line graphs. They represent time series data. However, bar graphs display a change in magnitude, and not in direction like line graphs. A horizontal bar graph consists of an x-axis, and a vertical bar graph consists of a y-axis. The numbers on the axes are known as the scales. Each bar is represents a numeric or categorical variable. Vertical bar graphs are best

used for the comparison of time series data and frequency distribution. Horizontal bar graphs are particularly useful when category labels are long; vertical bar graphs do not provide much space for text labels. There are a lot of varieties of bar graphs in common use. Horizontal or vertical bars (fully shaded rectangles) with the same width, drawn with their bases on the same horizontal or vertical line with equal gaps in between and lengths proportional to the magnitudes of the observations constitute a simple bar diagram. Subdivided bar diagram or component bar diagrams are used when the observations corresponding to the various categories have different components and it is felt that a comparison of the component parts is important. Here a simple bar diagram is first drawn with the length of the bars proportional to the totals of the component parts and then it is sub divided into parts of length proportional to the component magnitudes and each part given a different colour or shading. In Percentage bar diagrams, the component parts are expressed as the percentages of the total and a component bar diagram is drawn with all bars having equal length. Multiple bar diagrams are used for representing two or more interrelated data, for facilitating comparison. Deviation bar diagrams are usually used to represent net quantities like net profit, balance payable, deficit or excess etc. as the observations may be positive or negative, the base line is usually drawn in the middle of the paper horizontally and positive values are indicated by bars of proportional length, drawn above the horizontal line and negative values by bars of proportional length drawn below the horizontal line.

Horizontal and vertical bar graphs may exist in several forms. Double and group bar graphs, whether vertical or horizontal, are used to compare data about the same location or things. These graphs make it possible to compare several features at once. However, they can become puzzling if they contain too many sets of data. While double and group bar graphs are very popular, other types of less popular bar graphs include age pyramid graphs, stacked bar graphs, and dot graphs.

Line graphs

Line graphs are the most popular types of graphs, because they are simple to create and easy to understand. They organize and present data in a clear manner and show relationships between the data. They are used for personal, educational, and professional reasons. Particularly popular in the fields of science and statistics, they can also forecast the results of data that is not yet gathered. While line graphs and bar graphs share the same purpose, line graphs display a change in direction, while bar graphs display a change in magnitude. Line graphs are used to display the comparison between two variables which are plotted on the horizontal x- and vertical y-axes of a grid. The x-axis usually represents measures of time, while the y-axis usually represents percentage or measures of quantity. Therefore, line graphs are commonly used as time series graphs that show differences in direction.

A time plot is similar to a line graph. However, it always plots time on the x-axis. For instance, you can learn about the production of cars in the year 1960 by plotting the time variable (in this case, the months of the year) along the x-axis and the number of cars built in each month along the y-axis. After plotting, you draw a continuous line connecting the points on the grid. The result is a visual representation of the peaks and dips of car production throughout the months of 1960. You can similarly plot the number of cars produced in the years of 1961 and 1962 on the same graph. Thus, you can easily compare multiple relationships.

Histogram

A histogram is a bar chart or graph showing the frequency of occurrence of each value of the variable being analysed. In histogram, data are plotted as a series of rectangles. Class intervals are shown on the 'X-axis' and the frequencies on the 'Y-axis' if the classes are of equal width and frequency density (f/c) on 'Y-axis' if the classes are of un equal width. The height of each rectangle represents the frequency or frequency density of the class interval. Each rectangle is formed with the other so as to give a continuous picture. Such a graph is also called staircase or block diagram. However, we cannot construct a histogram for distribution with open-end classes.

Frequency Polygon and Frequency Curve

If we mark the midpoints of the top horizontal sides of the rectangles in a histogram and join them by a straight line, the figure so formed is called a Frequency Polygon. This is done under the assumption that the frequencies in a class interval are evenly distributed throughout the class. The area of the polygon is equal to the area of the histogram, because the area left outside is just equal to the area included in it. Another method of drawing frequency polygon is on the X axis draw the mid points and on the Y axis the frequency density (f/c) join the points by straight line to obtain frequency polygon. If the middle point of the upper boundaries of the rectangles of a histogram is corrected by a smooth freehand curve, then that diagram is called frequency curve. The curve should begin and end at the base line.

Ogives

The cumulative frequency gives the cumulative frequency of each of the class. The curve table is obtained by plotting cumulative frequencies is called a cumulative frequency curve or an ogive. There are two type of ogive namely: The 'less than ogive' and the 'greater than ogive'. In less than ogive method we start with the upper limits of the classes and go adding the frequencies. When these frequencies are plotted, we get a rising curve. In greater than ogive method, we start with the lower limits of the classes and from the total frequencies we subtract the frequency of each class. When these frequencies are plotted we get a declining curve.

Funnel Chart

Funnel charts can be used to illustrate stages in a process — usually sales processes. However, they could be used to show anything that's decreasing in size. For example, you can use this type of chart to show an order fulfillment process, a sales process from start to finish, the flow of information from top secret to unclassified or knowledge areas from general knowledge to expert knowledge. A funnel chart usually shows an overall decrease at each stage. When the stages in a funnel chart increase, it is called a pyramid graph. These charts are similar to pie charts, because the total area of the colored bars is 100%. For example, you might have one bar with an area of 40%, the next with 30%, then 22% and then 8%. Essentially, we can think of a funnel chart as a differently-shaped pie chart. The top of the funnel is called the head. It's also called the base. In this case, you can think of the "base" as being the base of a triangle. The lowest part of the chart is called the neck. We can create a funnel chart in Excel by typing your data into two columns, selecting all of the data and then by clicking funnel from Insert menu.

A funnel plot is not a synonym for funnel chart. A funnel plot is a scatter plot of the effect of a treatment vs. a measure of study precision or size. While funnel charts have bars showing stages, funnel plots have points for each data point. Essentially, the only thing they have in common is their overall shape.

Stemplot

A stemplot is like a histogram. They are both tools to help you visualize a data set. Stemplots show a little more information than a histogram and have been a common tool for displaying data sets since the 1970s. They are typically used when there is a medium amount of quantitative variables to analyze; Stemplots of more than 50 observations are unusual. The name "Stemplot" comes because there is one "stem" with the largest place-value digits to the left and one "leaf" to the right. A stem and leaf plot is a way to plot data where the data is split into stems (the largest digit) and leaves (the smallest digits). They were widely used before the advent of the personal computer, as they were a fast way to sketch data distributions by hand. They are used less frequently today, but you'll still see some here and there. The stem and leaf plot is used like a histogram; it allows you to compare data**.** While a histogram uses bars to represent amounts, the leaves of the stemplot represent amounts. A very long leaf means that "stem" has a large amount of data. The numbers are arranged by place value. The largest place-value digits are placed in the stem.

Conclusion

Since a diagram is a visualization of information, it can more effectively present data with less effort. Pictorial presentation of data is more attractive and evident than numbers and tables. More information can be effectively conveyed using a well-constructed diagram than a narrative-based process. This paper covers only a few diagrams and graphs which play a vital role in making statistics understand easy. The teaching of all branches of science can be enriched by the use of diagrams.

References

1. Agarwal B L, Basic Statistics, Revised Edition (2006), New Age International.

2. Guptha S C, Fundamentals of statistics, seventh edition (2017) Himalaya publishing house.

3. Franzblau L E, Chung K C. Graphs, tables, and figures in scientific publications: the good, the bad, and how not to be the latter. *J Hand Surg Am. (*2012); 37(3):591–596.

4. Ng K H, Peh W C. Preparing effective illustrations. Part 1: graphs. *Singapore Med J.* 2009;50(3):245.