**EXPERIMENT NO: 7**

**Aim: -** Statistical Graphics and its use in visualization.

**Theory:-**

**What is Statistical Graphics?**

A statistical graphic is data presented in a graphical format. A well-designed statistical graphic, also referred to as a chart or graph, consists of complex ideas communicated with clarity, precision, and efficiency. It gives its viewer the greatest number of ideas, in the shortest time, and in the smallest space, and with least possible clutter. It will also induce the viewer to think of substance, not techniques or methodology. It will provide coherence to large amounts of information by tying them together in a meaningful way, and it will encourage data comparisons of its different pieces by the eye. In general, a statistical graphic should possess the following qualities:

■ The objective and use of the graph should be obvious and apparent.

■ the graph type should be recognizable.

■ the graph type should help users understand the data more easily.

■ the data should be formatted and presented correctly.

■ the data should be formatted and presented for the using audience.

■ the graph should avoid distortions by telling the truth about the data.

**Use of Statistical Graphics:-**

■ Graphs are used for

— Simple comparisons of data.

— Presentation of changes over time.

— Statistical analysis.

— Illustration of proportions.

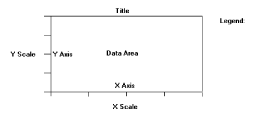
■ Reserve for material that is rich, complex, or difficult.

— Less than or equal to 3 numbers — use a sentence.

— For 4 to 20 numbers — use a table.

— More than 20 numbers — use a graph.

**Components of Statistical Graphics:**

Most statistical graphics have at least two axes, two scales, an area to present the data, a title, and sometimes a legend or key, as illustrated in figure given below. Pie charts are the exception to this general rule. Guidelines for graph components include the following.-

**Axes**

■ Values on an axis should increase as they move away from the origin.

■ Use the horizontal axis (X) to show time or cause of an event (the independent variable).

■ Use the vertical axis (Y) to show a caused effect (the dependent variable).

■ Provide additional axes when appropriate. For example, to show

**Scales and Scaling**

■ Place ticks to marks scales on the outside edge of each axis.

■ Employ a linear scale.

■ Mark scales at standard or customary intervals.

■ Create relatively square grid cells.

■ Start a numeric scale at zero (0).

■ Keep the number of digits in a scale to a minimum.

■ for scales containing decimals, put zeros in front of the decimal.

■ If data points are on the zero (0) line, offset the scale.

■ Display only a single scale on each axis.

■ for large data matrices, consider displaying duplicate axes.

■ Provide aids for scale interpretation.

■ Provide scaling consistency across two or more related graphics.

■ clearly label each axis in a left-to-right reading orientation.

**Proportion**

■ Provide accurate proportion of the displayed surfaces to the data they represent.

■ Provide proper proportion by

— Conforming to the shape of the data.

— Making the width greater than the height.

**Lines**

■ Data lines should be the heaviest.

■ Axes lines should be of medium weight.

— Extend the lines entirely around the graphic.

■ Grid lines should be very thin or absent.

— Let the user turn the grid on or off.

**Labeling**

A graph also needs a label for its title, the axes, its key or legend, and its data. First,

However, some general labeling guidelines follow.

■ Employ clear, detailed and thorough labeling.

■ Maintain a left-to-right reading orientation.

**Title**

■ Create a short, simple, clear, and distinctive title describing the purpose of the graphic.

■ Position the title above, centered, or left-aligned to the rectangle formed by the extended axes.

■ Spell it out fully, using a mixed-case or uppercase font.

**Axis Labels**

■ Center the X-axis label below its scale.

■ Center the Y-axis label above the scale or adjacent to the scale.

■ State units of measurement.

■ Provide information about the source of the data.

**Key or Legend**

■ Incorporate a key or legend for complicated graphs.

**Data Labels**

■ Permit data labels to be turned on or off.

**Aiding Interpretation of Numbers**

■ Display a grid on request.

■ Permit the viewer to click on a data point to display actual values.

■ Show numeric values automatically for each point or bar.

■ Permit the viewer to zoom in on an area of the graphic.

■ Permit the user to change the scale values.

■ Permit toggling between a graphic and a table

**Types of Statistical Graphics:-**

Statistical graphics take many forms. There are curve and line graphs, surface charts, scatter plots, bar charts, histograms, segmented or stacked bars, and pie charts.

1. **Curve and Line Graphs**

Curve and line graphs can be used to show relationships between sets of data defined by two continuous variables. They are especially useful showing data changes over time, being superior to other graphic methods for speed and accuracy in determining data trends. With a curve, a smoothed line summarizes the data relations. With a line, straight line segments connect the data plots. A line graph is illustrated below. This kind of graph implies a continuous function...

.****

**Uses**

■ Display data curves or lines that must be compared in a single graph.

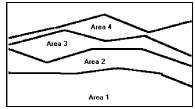
■ Display no more than four or five curves in a single graph.

■ Identify each curve or line with an adjacent label whenever possible.

■ if a legend must be included, order the legend to match the spatial ordering of the lines.

1. **Surface Charts**

If the data is being depicted by a curve or line represents all the parts of a whole, consider developing a surface or area chart, as illustrated in figure given below. In this kind of graph, the curves or lines are stacked above one another to indicate individual amounts or aggregated amounts. Each boundary height is determined by the height of the line below it, and the area between each line or curve is differently coded, usually by textures or shading. A surface chart is similar to a segmented bar chart.



**Uses**

■ Order the data categories so that

— The least variable is at the bottom, and the most variable at the top.

— The largest is at the bottom and the smallest at the top.

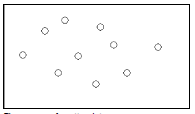
■ Use different texture or shading coding schemes to differentiate the areas below

Each curve or line.

■ Incorporate labels within the bands of data.

1. **Scatter Plots**

Scatter plots can be used to show relationships among individual data points in a two-dimensional array. A point is displayed on the plot where the X-axis and Y-axis variables intersect, as illustrated in figure below. Correlations and trends on scatter plots can be indicated by the superimposition of curves (thus combining the scatter plot with another kind of graphic display).

****

Two dimensions**.** Limit scatter plots to two dimensions.

Three-dimensional scatter plots, while possible, do not yield clear, unambiguous displays. Consistent intervals. Maintain consistent scale size intervals. Inconsistent spacing size between scale ticks on the two axes will distort the displayed data.

Distinguishable plots.Construct the plot points of distinguishable, equal-sized circles, squares, rectangles, or diamonds. These symbols may be filled in or empty. Color may also be used to designate the points. Keep in mind that, when using color, different colors can look different in size, and some people using the graphic may be colorblind...

**Uses**

■ Limit use to two-dimensional displays of data.

■ Maintain consistent scale size intervals.

■ Provide distinguishable, equal-sized plot points.

■ if there is more than one set of data on the plot, use different symbols for each data

Set’s points.

■ visually distinguish points of particular significance through a highlighting technique.

1. **Bar Charts**

Bar charts can be used to show a few differences between separate entities or to show differences in a variable at a few discrete intervals. They are useful for comparing or ranking a small number of values — no more than 10 or 12. A bar chart consists of a series of bars extending from a common origin or baseline, as illustrated in figure given below:-



**Uses**

■ Highlight bars representing important or critical data.

■ Provide a consistent ordering for related groups of bars.

■ Display a reference index if displayed data must be compared to a standard or critical value.

**Segmented or Stacked Bars**

If both the total measure of a value and its component portions are of interest, consider using segmented or stacked bars. These bars are similar to bar charts except that the bar is segmented into two or more pieces reflecting the component values, as illustrated in figure given below. In this way they are similar to surface graphs and pie charts. Design guidelines are similar to stacked bars, except for the following:



Data category ordering.To provide consistency, order the data categories in the same sequence for all bars. Order data categories to show least variable at bottom and most variable at top.

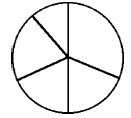
Large segments**.** Limit the number of segments to those that are big enough to be seen and labeled. If small segment components exist, group them into an “other” category.

Coding schemes.Use different texture or shading coding schemes. Ensure that the coding scheme chosen for each segment is visually distinguishable from all others. Place darker shades or colors toward the bottom or toward the left.

Labeling.Associate labels with bars and segments. Labels, with a left-to-right reading orientation, are preferable to legends. Do not place labels within segments, as they most often will not fit. Legends should only be used if space does not allow labels.

1. **Pie Charts**

Pie charts, a circle broken up into pie-shaped pieces, can be used to show an apportionment of a total into its component parts, as illustrated in figure given below Bar graphs, however, usually permit more accurate estimates of proportions.



Experts caution against the use of pie charts because

■ they provide no means of absolute measurement.

■ they cannot represent totals greater than 100 percent.

■ they can only represent a fixed point in time.

■ Human estimation of relationships is more accurate with linear than with

Angular representations.

1. **Flow Charts**

If the data to be displayed flows in a complex, yet sequential, process, consider using a flowchart to schematically represent it. Flowcharts can also be used to aid problem solving in which a solution can be reached by answering a series of questions. They are not useful when trade-offs must be made.

Order of steps.One logical ordering scheme is to follow a sequence of operations or processes from start to finish. Other potential ordering schemes include placing the most important decisions first or the decisions that can be made with the

Most certainty.

Orientation. Follow a left-to-right and top-to-bottom orientation.

Coding conventions.Follow existing shape coding conventions for the kinds of boxes being displayed. Adhere to standards and people’s expectations.

Arrows.Use arrows to indicate directional relations and sequential links.

**Conclusion**:-

Hence we have studied various types of Statistical Graphics.

We have used \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ type of Statistical Graphics for \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ applications.

The advantages for using Statistical Graphics for our application are:-

1)

2)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Program Execution**  **(7)** | **Documentation**  **(2)** | **Punctuality**  **(2)** | **Viva**  **(4)** | **Experiment**  **Marks**  **(15)** | **Teacher**  **Signature**  **with date** |
|  |  |  |  |  |  |