PRESENTATION OF DATA

Shahbaz Baig

Abstract: Most scientific experiments are conducted in an attempt to answer some specific questions and generally result in the collection of data in the form of batches of numbers, usually referred to as sample. To extract information from the sample, there is need to organize and summarize the collected data. There are many ways of describing a sample. Commonly, we use either a graph or a small set of numbers which summarize some properties of the sample such as its center and spread. Arrangement of data according to the values of a variable characteristic is called a distribution. When the defining variable is expressed in terms of location, we get a spatial or geographical distribution.

Key words: Table, Graph, Chart, Frequency Distribution, Histogram, Polygraph, Whiskers Plot.

Introduction

Most scientific experiments are conducted in an attempt to answer some specific questions and generally result in the collection of data in the form of batches of numbers, usually referred to as sample. To extract information from the sample, there is need to organize and summarize the collected data. There are many ways of describing a sample. Commonly, we use either a graph or a small set of numbers which summarize some properties of the sample such as its center and spread[1].

Classification

The term classification is defined as the process of dividing a set of observations or objects into classes or groups in such a way that

i Observations or objects in the same class or group are similar

ii Observations or objects in each class or group are similar to observations or objects in other class or group.

Classification is thus the sorting of data into homogenous classes or groups according to their being alike or not.

When the data are sorted according to one criterion only it is called a simple classification or a one-way classification. Classification is called a two-way classification when the data are sorted according to two criteria.

A manifold classification or cross-classification is made according to several criteria.

Data may also be classified according to qualitative, temporal and geographical characteristic.

Article Citation: Baig S, Presentation of Data. Indep Rev Jan-Jun 2017;19(1-6): 61-75.

Date received: 18/03/2017
Date Accepted: 08/05/2017

Correspondence Address:
Dr. Shahbaz Baig, MBBS, MPH
Assistant Professor of Community Medicine
Independent Medical College, Faisalabad.

1. Dr, Shahbaz Baig, MBBS, MPH
Assitant Professor Community Medicine,
Independent Medical College, Faisalabad.

Arrangement of data according to the values of a variable characteristic is called a distribution. When the defining variable is expressed in terms of location, we get a spatial or geographical distribution. Temporal arrangement of values is referred to as a time series[2].

Aims of classification

The main aims of classification are:

i. To reduce the large sets of data to an easily understood summary;
ii. To display the points of similarity and dissimilarity;
iii. To save mental strain by eliminating unnecessary details;
iv. To reflect the important aspects of the data; and
v. To prepare the ground for comparison and inference[2].

Basic principles of classification

While classifying large sets of data, the following points should be taken onto consideration.

i. The classes are categories into which the data are to be divided, should mutually exclusive and no overlap should be exist between successive classes. In other words, classes should be arranged so that each observation or object can be placed in one and only one class.
ii. The classes or categories should be all inclusive. All inclusive classes are classes that include all the data.
iii. As far as possible, the conventional classification procedure should be adopted.
iv. The classification procedure should not be so elaborate as to concentrate all the data in one or two classes[2].

PRESENTATION OF DATA

Statistical data can be presented in two main groups:

I. Tables
II. Graphical Presentation
a) Diagrams
b) Graph

TABLE

A table is a systematic arrangement of data into vertical columns and horizontal rows. The process of arranging the data into rows and columns is called tabulation.

Depending upon the number of characteristics involved the tabulation may be classified as;

- Simple
- Double
- Complex

General rules of tabulation

There are certain rules which should be followed in tabulation.

- A table should be simple, easy to understand. There should be no need to go through footnotes or explanations.
- If the observations are large in number they can be broken into different groups and more then one table can be prepared.
- Proper and clear headings for columns and rows should be used.
- Thick lines should be used to separate the data under big classes and thin lines to separate the sub-classes of data.
- The unite of measurement under each heading and sub-heading must be indicated.
- Totals of columns should be preferably at the bottom of the table and the totals of rows should be at the extreme right side of the table.

It is not possible to make right rules for tabulation. In general tables should suit the needs and requirements of an investigation3.

Types of Tables

Statistical tables classified according of purpose, are of two types, General purpose (primary) tables.

The general purpose tables are large in size, are extensive with vast coverage and are constructed for reference purses. Specific purpose (derived or text) tables.The specific purpose tables are simpler in structure and deal with one or two criteria of classification only. Such tables are used to analyses or to assist in analyzing data.

CONSTRUCTION OF TABLES

Following are the parts of table out of which first four are main part.

Title

A title is the heading at top of the table. The title should be brief and self explanatory. It describes the contents of the table.

Column Captions and Box-head

The heading for different columns are called column captions and this part of column captions is called Box-head. The column captions should be brief, clear and arranged in order of importance.

Row Captions and Stub

The headings for different rows are called row captions and the part of the table containing row captions is called Stub. Row captions should be brief, clear and arranged in order of importance.

Body of the Table

The entries in different cells of columns and rows in a table is called body of the table. It is the main part of the table. The data may be arranged qualitatively, quantitatively, chronologically, geographically or alphabetically.

Source Note

Source notes are given at the end of the table which indicate the compiling agency, publication, the data and page of the publication.

Spacing and rulings

To enhance the effectiveness of a table, spacing and ruling is used. It is also used to separate certain items in the table. Thick or double lines or single lines are used to separate row captions and column captions. To indicate no entry in a cell of the table, dots (…) or dashes (---) are used. Zeroes are not used in a table for this purpose.

Prefatory Notes and Footnotes

The prefatory note is given after the title of the table and the footnotes are given at the bottom of the table. Both are used to explain the contents of the table. The footnotes are usually indicated by[3].

FREQUENCY DISTRIBUTION

The organization of a set of data in a table showing the distribution of the data into classes or groups together with the number of observations in each class or group is called a frequency distribution. The number of observations falling in a particular class is referred to as the class frequency or simply frequency and is denoted by f. Data presentation in the form of a frequency distribution are also called grouped data while the data in the original form are referred to as ungrouped data. The data are said to be arranged in array when arranged in ascending in descending order magnitude. The purpose of a frequency distribution is to produce a meaningful pattern for the overall distribution of the data from which conclusions can be drawn. A fairly common frequency pattern is rising to a peak and then declining. In terms of its construction, each class or group has lower and upper limits, lower and upper boundaries, an interval and a middle value[2].

Class limits

The class-limits are defined as the number or the values of the variable which describe the classes; the smaller number is the lower class limit and the larger number is the upper class limit.

- Class limits should be well defined and there should be no overloading. In other words, the limits should be inclusive, i.e. the values corresponding exactly to lower limit or the upper limit be include in that class.
- The class-limits are therefore selected in such a way that they have a same number in significant places as the recorded values.
- Suppose the data are records to the nearest integers. Then an appropriate method for defining the class limits without overlapping, for example, may be 10-14,15-19, 20-24, etc. The class limits may be defined as 10.0-14.9, 15.0-19.9, 20.0-24.9, etc.
- When the data are recorded to nearest tenth of an integer.
- Sometimes, a class has either no lower class limit or no upper class-limit. Such a class is called an open-end class. The open-end classes, if, possible, should avoided as they are a hindrance in performing certain calculations. A class indicated as 10-15 will include 10 but not 15, i.e. 10≤X<152.

CLASS BOUNDARIES

The class boundaries are the precise number which separate one class from another. The selections of these numbers remove the difficulty, if any, in knowing the class to which a particular value should be assigned.

A class-boundary is located midway between the upper limit of a class and the lower limit of next higher class, e.g. 9.5-14.5, 14.5-19.5, 19.5-24.5, or 9.95-14.95, 14.95-19.95, etc.

The class-boundaries are thus always defined more precisely than the level of measurements being used so that the possibility of any observation falling exactly on the boundary is avoided. That is why the class boundaries carry one more decimal place than the class limits or the observed values.

The upper class boundary of a class coincides with the lower class boundary of the next class.

Class Mark. A class mark also called class midpoint, is that number which divides each into two equal parts.

In practice, it is obtained by dividing either the sum of the lower and upper limits of a class, or the sum of the lower and upper boundaries of the class by 2 but in a few cases, it does not hold, particularly in modern practice of age grouping.

For purposes of calculations, the frequency in a particular class is assumed to have the same values as the class-mark or midpoint. This assumption may introduce an error, called the grouping error, but statistical experience has shown that such errors usually tend to counterbalance over the entire distribution.

CLASSES (WEIGHT)	CLASS-BOUNDARIES	MID-POINT OR CLASS MARKS	TALLY	FRE-QUENCY
65-84	64.5-84.5	74.5	~~IIII~~ IIII	9
85-104	84.5-104.5	94.5	~~IIII~~	10
105-124	104.5-124.5	114.5	~~IIII~~ ~~IIII~~ ~~IIII~~ II	17
125-144	124.5-144.5	134.5	~~IIII~~ ~~IIII~~	10
145-164	144.5-164.5	154.5	~~IIII~~	5
165-184	164.5-184.5	174.5	IIII	4
185-204	184.5-204.4	194.5	~~IIII~~	5
Total	.....	.....	.....	60

The grouping error may also be minimized by selecting a class (group) in such a way that its midpoint corresponds to the mean of the mean of the observed values falling in that class[2].

Class Width or Interval

The class-width or interval of a class is equal to the difference between the class boundaries.

It may also be obtained by finding the difference either between two successive lower class limits, or between two successive class marks.

The lower limit of a class should not be subtracted from its upper limit to get the class interval.

An equal class interval, usually denoted by h or c, facilitates the calculations of statistical constants such as the mean, standard deviation, moments, etc. That is why in practice, it is desirable to have equal class-intervals.

But in some types of economic and medical data, it is wise to use unequal class-intervals on account of greater concentration of measurements in certain classes. Such class intervals usually become uniform when logarithms of class are taken.

It should be noted that some people use the terms "class" and "class-interval" interchangeably and the width of the class is referred to as the size or length of the class-interval2.

Constructing a Grouped Frequency Distribution

The following are some basic rules that should be kept in mind when constructing a grouped frequency distribution:

Decide number of classes

There are no hard and fast rules for deciding on the number of classes which actually depends on the size of data. Statistical experience tells us that no less than 5 and no more than 20 classes are generally used. Use of too many classes will defeat the purpose of condensation and too few will result in too much loss of information.

H.A. Sturges has proposed an empirical rule for determining the number of classes into which a set of observations should be grouped. The rule is

$$k = 1 + 3.3 \log N$$

Where k denotes the number of classes and

N is the total number of observations. For example, if there are 100 observations, then by applying sturges rule we should have

k = 1+3.63 (2.0000) = 7.6, i.e. 8 classes

Thus eight classes are required but this rule is rarely used inn practice.

Determine the range of variation in the data, i.e.

The difference between the largest and the smallest values in the data.

Divide the range of variation by the number of classes:
To determine the approximate width or size of the equal class-interval. In case of factional results, the next higher whole number is usually taken as the size or width of class-interval.

If equal class-interval are inconvenient or be undesirable, then classes of unequal size are used. But in practice, intervals that are multiple of 5 or 10, are commonly used as people can understand them more readily.

Decide where to locate the class-limit: Of the lowest class and then the lower class boundary.

The lower class usually stars with the smallest data value or a number less than it. It is better if it is a multiple of class-interval.

Find the upper class boundary by adding the width of the class-interval to the lower class-boundary and write dawn the upper class limits too.

The open-end classes, i.e. classes with the lowermost or uppermost class boundary unknown, should be avoided if possible.

Determine the remaining class-limits: And class boundaries by adding the class-interval repeatedly.

The lowest class should be placed at the top and the rest should follow according to size.

In some cases, the highest class is placed at the top.

Distribute the data into the appropriate classes.
This is best done by using a Tally-Column where values are tabulated against appropriate classes by merely making short bars or tally marks to represent them.

It is customary for convenience in counting to place the first four bars vertically and the fifth one diagonally and to leave a space.

The number of tallies is then written in the frequency column.

The tally column usually omitted in the final presentation of the frequency distribution.

But in case of small number of values, the actual values should be shown against each class to mitigate chances of error.

Finally, total the frequency column: To see that all the data have been accounted for. These rules are applied to group raw data which arte assumed to be continuous.

In case of discrete data which carry only integral values, the concept of a class boundary is unrealistic as there can be no points where, the adjoining classes meet.

In spite of this logical difficulty, when the discrete data are sufficiently large, they are treated for convenience of calculations as continuous and hence are grouped in the same way as the continuous data.[2]

Example

Make a grouped frequency distribution from the following data, relating to the weight recorded to the nearest grams of 60 apples picked out at random from a consignment.

By scanning the data, we find that the largest weight is 204 grams and the smallest weight is 68 grams so that the Range is 204-68= 136 grams.

Suppose we decide to take 7 classes of equal size. Then size or width of the equal class interval would be 136/7 = 19.47.

But we take h=20, the next integral value higher than 19.47 to facilitate the numerical work.

Let us decide to locate the lower limit of the lowest class at 65. With this choice, the class limits will be 65-84, 85-104, 105-124,…….
the class boundaries become 64.5-84.5, 84.5-10.5,104.5-124.5,…….,and the class-marks are 74.5, 94.5, 114.5…. The grouped frequency distribution is then constructed as follows:

By listing the actual values

Frequency Distribution Of

Weights Of 60 Apples

This table is sometimes known as an entry table. The values against each class may be arranged in an array.
By Using A Tally – Column:

106	107	76	82	109	107
	115	93	187	9	5
123					
	125	111	92	86	70
	126	68	130	129	139
	119	115	128	100	186
	84	99	113	204	111
	141	136	123	90	115
98	110	78	185	162	178
	140	152	173	1 4	6
158					
	194	148	90	107	181
	131	75	184	104	110
	80	118	82		

Frequency Distribution Of Weights Of 60 Apples

Weight	Entries	Frequency
65-84	76, 82, 70, 68, 84, 78, 75, 80, 82,	9
85 – 104	93, 95, 92, 86, 100, 99, 90, 98, 90, 104	10
105 - 124	106, 107, 109, 107, 115, 123, 111, 119,115, 113, 111, 123, 115, 110, 107, 110,118	17
125 - 144	125, 126, 130, 129, 139, 128, 141136,140, 131	10
145 - 164	162, 152, 146, 158, 148	5
165 - 184	178, 173, 181, 184	4
185 - 204	187, 186, 204, 185, 194	5

CUMULATIVE FREQUENCY DISTRIBUTION

The total frequency of a variable from its one end to a certain value (usually upper class boundary in grouped data), called the base, is known as the cumulative frequency, less than or more than the base of the variable.

A table that shows the cumulative frequencies, is called a cumulative frequency distribution. The cumulative frequency of the last class is the sum of all frequencies in the distribution.

If the cumulation process is from the lowest value to the highest, it is referred to as "a less than" type cumulative frequency distribution. Where the frequencies are cumulated from the highest value to the lowest value, it is called a "more than" type cumulative frequency.

Relative Frequencies

If the class frequencies against various classes are divided by the total frequency, we get the relative frequencies which always add to one. The class frequencies may also be expressed as percentages, the total of which would be 100. A percentage cumulative distribution is useful to read off the percentage of values falling between certain specified values.

Weight(grams)	Cumulative Frequency (F)
Less than 64.5	0
Less than 84.5	9
Less than 104.5	19
Less than 124.5	36
Less than 144.5	46
Less than 164.5	51
Less than 184.5	55
Less than 204.5	66

A "less then" type cumulative frequency distribution is shown below:
A "more than" type cumulative frequency distribution is given below:

STEM – AND – LEAF DISPLAY

Introduction

A clear disadvantages of using a frequency table is that the identity of individual observation is lost in grouping process. To overcome this drawback "John Tokey" 1977 introduced a technique known as the stem and leaf display. This technique offers a quick and novel way for simultaneously sorting and displaying data sets where each number in the data set is divided into two parts a stem and a leaf.

- A stem is the leading digits of each no and is used in sorting.
- A leaf is the rest of the number or the trailing digits and shown in display

A vertical line separates the leaf/leaves from the stem.
Example:- The number 243 could be split two ways.

Weight (grams)	Cumulative Frequency (CF)
More than 64.5	60
More than 84.5	51
More than 104.5	41
More than 124.5	24
More than 144.5	14
More than 164.5	9
More than 184.5	5
More than 204.5	0

Leading digit = 100	Trailing digit = 1	Leading digit = 10	Trailing digit = 1
2 Stem	43 Leaf	24 Stem	3 Leaf

All possible stems are arrangede in order from the smallest to the largest and placed on the left hand side of the line.

The stem – and – leaf display is a useful step for listing the data in an array. Leaves are associated with the stem to know the numbers. The stem – and - leaf table provides a useful description of data set and can easily be converted to a frequency table.

9.0	10.2	11.3	12.1	10.7	13.8	10.8
11.6	13.6	16.4	11.0	15.8	9.3	13.7
11.7	11.0	8.0	12.0	11.5	9.7	11.6
10.1	14.1	10.0	9.9	13.4	15.7	11.5
12.3	9.8	13.0	9.1	8.3	12.9	14.0
10.5	13.2	10.5	10.6	12.5	15.1	12.8
10.4	11.2	9.3	11.7	17.7	13.9	16.9
13.4	11.8	16.8	14.3	11.8	9.6	11.9
8.7	14.7	10.9	17.9	11.5	14.7	15.9
11.8	10.6	12.6	12.6	15.7	14.9	9.9

$$Range\begin{cases}\text{maximum value} = 17.9\\ \text{Minimum value} = 8.0\end{cases}$$

Stem	Leaf (decimals)	LEAF (ordered)
8	703	037
9	03193769	013367899
10	15462059678	01245566789
11	678820370585659	002355566778889
12	36601598	01356689
13	42604397	02446789
14	712790	012779
15	87719	17789
16	489	489
17	97	97

It is a common practice arranges the trailing digits in each row from smallest to highest2.

Make a stem and leaf table for the following data.

GRAPHIC REPRENTATION

Statistical tables contain data in the form of figures. But numerical are usually not attractive and some people find it difficult to get a clear picture from the numerical data. A more attractive method of presenting the data is to make good looking diagrams and graphs. There are very large number of graphs, pictures and diagrams which are used to represent data. Good looking diagrams and pictures leave an everlasting impression on the mind of the observer.
The modern advertising is mainly based on picture of various type to attract the consumers.

The visual representation of data is known as graphical representation. Such visual representation can be divided in to:

A. Diagrams
B. Graphs

The basic difference between a graph and a diagram is that a GRAPH is a representation of data by a continuous curve while, DIAGRAM is any other visual representation.[4]

DIAGRAMS

When ever a comparison of the same type of data at different places is to be made, diagram will be the best way to do that. Diagrammatical representation has several advantages over tabulation. Beautifully and neatly constructed diagrams are more attractive than simple figures. Comparison is made easier with diagrams. Diagrams are further divided in to following groups2.

Types of Diagrams

a): Bar Charts:
   i. Simple bar charts
   ii. Multiple bar charts
   iii. Component bar charts

b): Pie Diagram

c): Pictogram
d): Scatter Diagram
e): The Box and the whisker

Bar Charts

A graph drawn using rectangular bars to show how large each value is.

The bars can be horizontal or vertical[5].

Bar charts are useful methods of presenting data by the length of bars.

They are frequently used because they are better retained in the memory.

Have powerful impact

Used as a tool for comparing mutually exclusive discrete data

- Simple bar charts
- Multiple bar charts
- Component bar charts[2]

Simple bar charts

It consists of simple charts.

a. Bar charts (vertical / Horizontal) are most useful, widely used popular, easy way of expressing statistical data of nominal or ordinal variable .
b. The categories / variable are usually indicated on the X – axis (abscissa) and frequencies are indicated on Y-axis (ordinate).
c. The length of bar is proportional to the magnitude of the variable.
d. The bars are separated form each other.
e. Each axis is labeled.
f. The figure is numbered and fitted.

OR

A seprat horizantal or vertical bar is drawn for each category, its length being perpotional to the frequency in that category. The bars are seprats by small gapes to indecate that the data categoricalll for discrete[6].

Example:- The data of Mayo Hospital blood band suggests that

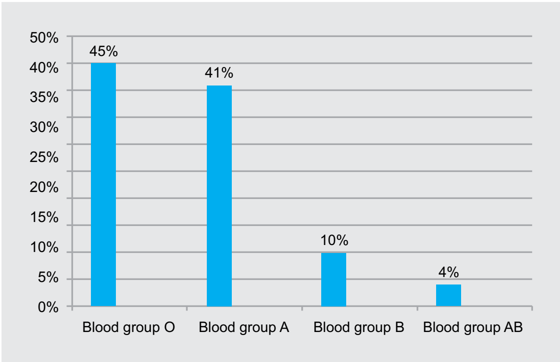
45 % ---------- blood group O
41 % ---------- blood group A
10 % ---------- blood group B
04 % ---------- blood group AB

Multiple Bar Chart

A multiple bar chart shows two or more characteristics corresponding to the values of a common variable in the form of grouped bars whose lengths are proportional to the values of the characteristics and each of which is shaded differently for identification.[7]

Example

Draw a multiple bar chart to represent the male population of four divisions of Punjab in 1961, 1963, and 1970.

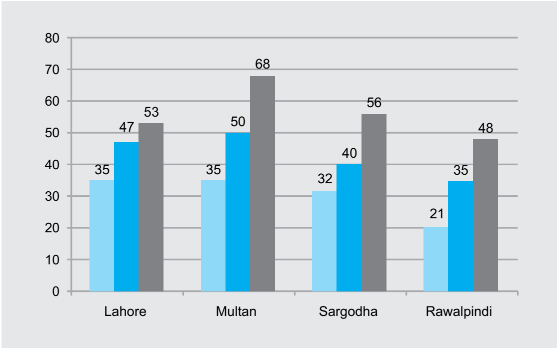


Component Bar Chart

In component bar chart each bar is divided into two or more sections. The length of the bar represents the total and various sections represent the components of total.

Example

Draw a sub-divided bar chart to represent the male and female population of four divisions of Punjab in 1961.



Division	Male	Female	Total
Lahore	35	30	65
Multan	35	31	66
Sargodha	32	28	60
Rawalpindi	21	19	40

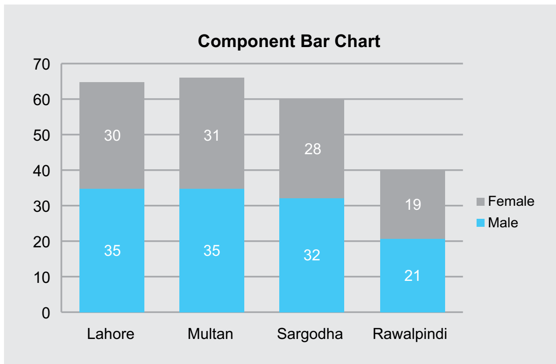
PIE DIAGRAM

A pie diagram consists of a circle divided into sectors whose areas are proportional to the various parts into which the whole quantity is divided. The proportion that each component part bears to the whole quantity will be the corresponding proportion of 360 degree.

A pie chart is a simple, easily understood chart in which the size of the "slices" or wedges shows the proportional contribution of each component part. Pie charts are useful for showing the proportions of a single variable's frequency distribution.

Given current technology, pie charts are almost always generated by computer rather than drawn by hand. But the default settings of many computer programs differ from recommended epidemiologic practice. Many computer programs allow one or more slices to "explode" or be pulled out of the pie. In general, this technique should be limited to situations when you want to place special emphasis on one wedge, particularly when additional detail is provided about that wedge .

Division	Male (000)		
	1961	1963	1970
Lahore	35	47	53
Multan	35	50	68
Sargodha	32	40	56
Rawalpindi	21	35	48



Multiple pie charts are occasionally used in place of a 100% component bar chart, that is, to display differences in proportional distributions. In some figures the size of each pie is proportional to the number of observations, but in others the pies are the same size despite representing different numbers of observations.

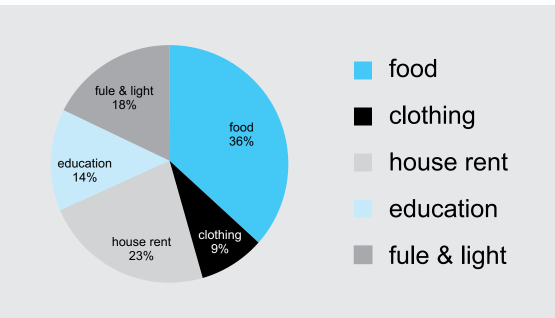
More About Constructing Pie Charts.
Conventionally, pie charts begin at
12 o'clock.
The wedges should be labeled and arrangedfrom largest to small proceeding clockwise, although the "other" or "unknown" may be last.

Shading may be used to distinguish between slices but is not always necessary.

Because the eye cannot accurately gauge the area of the slices, the chart should indicate what percentage each slice represents either inside or near each slice8.

Example: Represent the following data through a pie diagram:

Items of expenditure	Amount	Angle
Food	4000	131
Clothing	1000	32
House rent	2500	82
Education	1500	49
Fuel and Light	2000	66



Pictogram

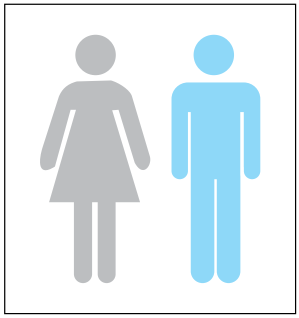
These consist of pictures or small symbolic figures representing the statistical data. It is the presentation of data for easy understanding of sweet persons (common man).
It is frequently used for quantization data.

Scatter Diagram

Scatter diagram is used to see the relationship between variables.

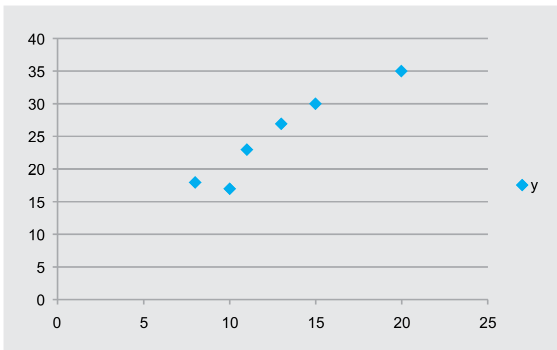
Example:

X	Y
8	18
1	17
11	23
13	27
15	30
20	35



Types of Graphs

Graph of time series or Historiogram



Histogram

Frequency polygon & Frequency curve
Cumulative Frequency polygon or Ogive
Percentage cumulative frequency polygon

Historiogram

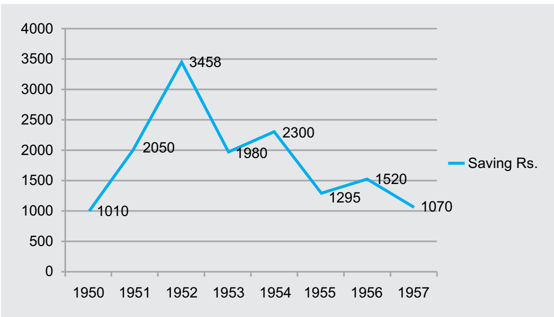
Historiogram is constructed by taking

1. Time along X-axis and
2. Value of the variable along Y-axis
3. Plot the points and
4. Connect the points by straight line segments to get the historiogram[9].

Example:

The data represent the records of a company's savings over the years. Construct a time series plot to represent it

Year	1950	1951	1952	1953	1954	1955	1956	1957
Saving Rs.	1010	2050	3458	1980	2300	1295	1520	1070



HISTOGRAM

A Histogram consists of a set of adjacent rectangles whose bases are marked off by

- Class boundaries along the X-axis
- Frequency along Y-axis
- Draw rectangles whose height are proportional to the frequencies with respective classes[10].

EXAMPLE

Frequency Polygon

- Frequency polygon is a graphic from of a frequency distribution,
- Take midpoints along X-axis
- Class frequency along Y-axis
- Plot the points and then connected by sraight line segments
- And extra class midpoint at both ends of the with zero frequency so that the polygon does form a closed figure with the horizontal axis[11].

(Polygon is a closed figure having many sides)

Frequency Curve

Frequency curve is constructed by taking the midpoints along X-axis and class frequency along Y-axis. Points are plotted and are then connected by free hand curve 10.

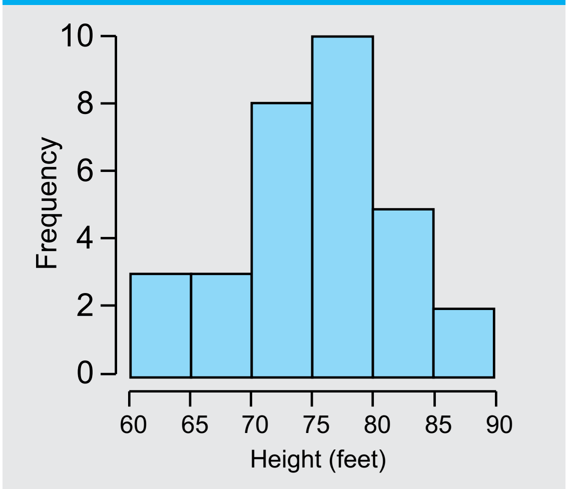
Class boundaries	Freq.
60-65	3
65-70	3
70-75	8
75-80	10
80-85	5
85-90	2

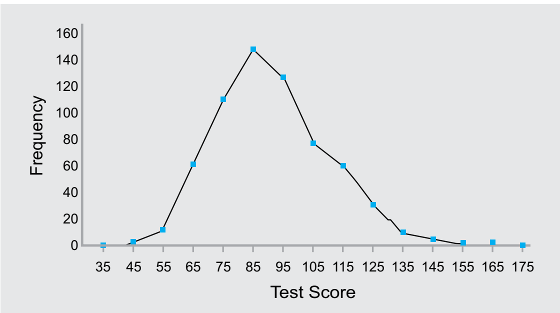
CUMULATIVE FREQUENCY POLYGON OR OGIVE

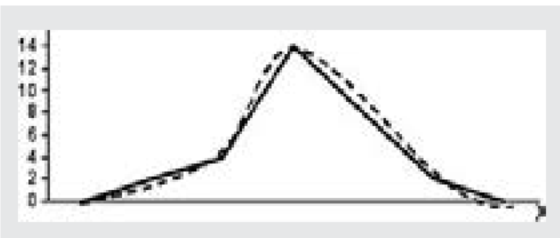
A cumulative frequency polygon known as ogive is a graph obtained by

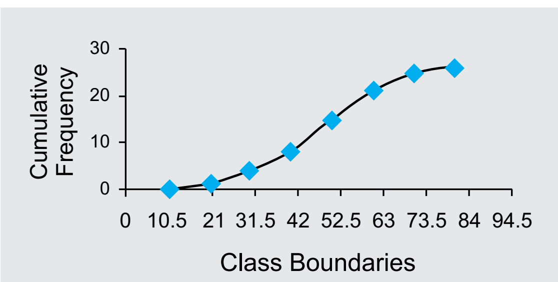
- On X-axis are the values of the variables and on Y axis cumulative frequencies.
- To construct "OGIVE" an ordinary frequency distribution table has to be converted into a cumulative frequency table. We place a point at upper class boundary of each class interval.
- Each point represents the cumulative frequency.
- OGIVE is computed by connecting all the points by a free hand curve.
- It is useful to compare two sets of data.
- Percentile may be obtained from an OGIVE.
- Shape of OGIVE somewhat resembles English letter "S" 12.

HEIGHT OF STUDENTS





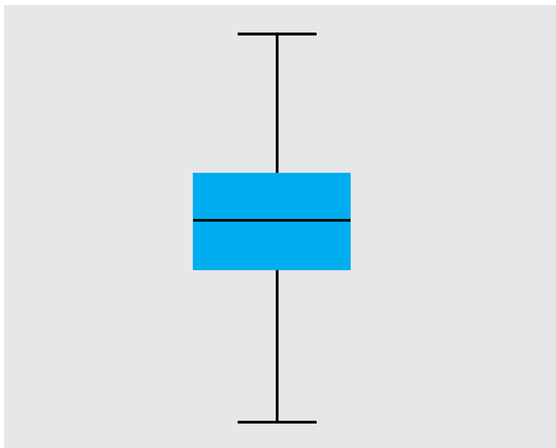




THE BOX PLOTS

The Box Plots, which are graphically very simple, are based on the Median, a measure of location and the Integrative. Range (IQR), a measure of data's variability. They are informative and effective for comparing two or more data sets / distributions.

A box plot is constructed by drawing a rectangle (he box) with the ends (called the hinges) drawn at the lower and upper quartiles (Q1 Q3). The median of the data is shown by the line dividing the box . The straight lines (called the whiskers) are dawn from each hinge to the most extreme observation. The entire graph is called a Box and Whiskers plot. If one whisker is longer, the distribution of data is skewed in the direction of the longer whisker. The box plot given below represents the distribution of examination marks given.



When two or more distributions are to be compared by drawing box plots, the scale of measurement is usually plotted vertically. Sometimes, two sets of limits, called inner fences and outer fences are also used13.

REFFERENCE

1. Singh, Simon (2000). The code book : the science of secrecy from ancient Egypt to quantum cryptography (1st Anchor Books ed.). New York: Anchor Books. ISBN 0385495323.^ Ibrahim A. Al-Kadi
2. (STATISTICAL METHODS & DATA ANALYSIS BY FAQIR MUHAMMAD) (Shair Muhammad Chudhary) (INTRODUCTION STATISTICAL THEORY BY SHAIR MUHAMMAS)
3. (A Text Book ofStatistics ,punjab text board book)
4. (BASIC STATISTICS BY KIANI)
5. Pierce, Rod. "Definition of Bar Graph" Math Is Fun. Ed. Rod Pierce. 13 Aug 2011. 27 Feb 2012 <http://www.mathsisfun.com/definitions/bar-graph.html
6. (MEDICAL STATISTICS AT A GLANCE)
7. Biostatistics for all Brigadier khan I A
8. Displaying Public Health Data Page 4-51principles Of Epidemiology In Public Health Practice Third Edition An Introduction To Applied Epidemiology And Biostatistics U.S. Department Of Health And Human Services Centers For Disease Control And Prevention (Cdc) Office Of Workforce And Career DevelopmentAtlanta, Ga 30333
9. Hanif M Biostatistics
10. Glaser AN.High yield Biostatistics 3rd edition.Lipi cot williams and wilkin.
11. Rao KV.Biostatistics; a manual of statistical methodes for use in health, nutrition and anthropology 2nd edition .Jaypee brothers medical publishers new Dehli India.
12. Lecture notes by Dr Anayat Khan HOD state department UAF