

A COURSE GUIDE TO GEOGRAPHICAL STATISTICS & QUANTITATIVE METHODS

Nagubwagye Abel
NATIONAL TEACHERS COLLEGE
MUNI
August 2019

[Document
subtitle]

STATISTICS

1.APPLICATION OF STATISTICS IN GEOGRAPHY

Statistics refers to a scientific and systematic methods of collecting, recording, summarizing, analyzing and representation of numerical data in precise manner.

Or

The study of methods of collecting, recording, summarizing, analyzing and presentation of data in precise manner by using numbers

Or

A science of observing, collecting, recording, summarizing, analyzing and presentation of data in precise manner by using numbers.

Numerical data understood as a body of information which given in numbers. Or Exact numerical facts or figures collected systematically and arranged for a certain purpose.

1.1 NATURE OF DATA.

1.2 Statistical data according to their varied nature

Statistical data according to their varied nature include the following: -

Discrete data.

It is a form of statistical data for variables whose values expressed or given in whole numbers. i.e. The data is for cases which do not exist in fractions. For instance; the data for the number of people which can be given as 102 people who can not be divided into either decimal or fractions.

Continuous data.

The data for the variables whose values can be expressed in fraction or decimals. In this type of data, any value within the range can be given. For instance; the data for temperature, rainfall, pressure, distance, growth rate, and other cases which also reflect the same. They are presented in continuity manner of fraction or decimals

Individual data.

The set of data which provides specific value to every item in a sample given. For instance; Juma has weight of 47 kg. They consider every item as an important entity and singly presented

Grouped data.

It is a form of data which gives values in range or classes. This type of data is of no precise as exact figures are quoted but values range in groups. The classic example of the grouped data is that of population distribution by age and sex which may appear as follows:-

AGE	FEMALES	MALES
0-9	14,897	14,567
10-19	15,432	14,329
20 – 29	17,987	13,098
30 – 39	16,876	17,654

Statistical data according to scale of measurements

This aspect is considerably on how the values of statistical data are given. The scale of measurement include the following.

Nominal data

The type of data according to scale of measurement of which the values are given according to the name of items in a given sample. e.g. 10 apples, 5 oranges, 7 mangoes, 5 banana and 2 cherish.

Ordinal data

The data of which the values are given in an order of magnitude of observation in such a way the numbers indicate the rank order among objects. i.e. the values are commonly given in either ascending or descending order e.g. 91, 82, 79, 74, 68, 67, 58, 54 and 49.

The interval data

The data of which values are given in range at regular distance by being grouped. e.g. The data for population distribution by age and sex expressed in interval scale.

Ratio data

The data of which the values given show the number of times items of has relatively to another e.g. 1:3, 2:5, 3:7. e.t.c.

VARIABLES

Variable is an attribute that has values of which fluctuate under a given condition . For instance; production is a considerable variable as whose values change under conditions of policies lie; climate, technology, marketability and other which may make the same.

Variables are considerably varied and are classified into dependent and independent variables.

Dependent variable

Dependent variable is the one whose values fluctuate due to the force of another variable. i.e. the variable whose values change irregularly as controlled by another variable. For instance; production is one among the most pronounced variables as changes due to the force of other variables like climate, level of technology applied, demand of the products produced, and others which might cause it to change.

1.2 CLASSIFICATION OF STATISTICS.

Statistics being the scientific and systematic methods dealing with numerical facts is broadly categorized into two depending on how data handled. The main broad categories include; descriptive and inferential statistics.

Descriptive Statistics

Descriptive statistics deal with recording, summarization, analyzing and presentation of numerical facts that have been actually collected. The actual collection of data can be like to population by conducting census.

Inferential statistics

Inferential statistics deal with recording, summarization, analyzing and presentation of numerical facts that have been handled by quantifying the uncertainties through prediction e.g. the likely harvest output in the next year or season.

STATISTICAL DATA

As already pointed out, statistical data are understood as the exact numerical facts or figures collected systematically and arranged for a certain purpose or body of information which is usually treated in numerical values.

Statistical data assessed being extremely varied and thus recognized be of different types. The categories of statistical data recognized with regards to their derived sources, varied nature and scale of measurements.

Statistical data according to their varied sources

Data by sources classified into two and include **primary** and secondary **data**.

Primary data

These are the numerical facts collected from the field or handled for the first time. i.e. They are the first hand or original information. The data are not available in the existing sources like books. Primary statistical data are handled by the techniques of interview, the use of questionnaires, observation, counting, measurements and other methods.

Secondary data

These are the numerical facts derived from the stored sources. The data were compiled by other people who carried out research. The sources of this type of data include; text books, reference books, magazines, maps, video tapes, audio tapes, and other sources which deliver the same.

Independent variable

Independent variable is the one whose values change on its own without being influenced by another variable. i.e. the variable whose values change steadily and regularly e.g. distance.

1.3 SOURCES OF STATISTICAL DATA

The sources of statistical data are simply the techniques employed to gather the numerical facts. These are broadly two and include; the numerical facts. These are broadly two and include; primary and secondary sources. Some of the primary techniques (sources) providing statistical data include the following:-

- Interview method
- Questionnaire
- Scheduling
- Field observation method
- Literature review

Interview method

The technique of interview involves the collection of data through the asking of questions verbally by researcher to a respondent.

Or

Is a verbal interaction between an interviewer and interviewee designed to list the information, news, opinion and feelings they have on their own. Generally an interview is an oral organization of questions asked to respondents by a researcher.

Questionnaire method

Questionnaire is a set of research questions printed on a piece of paper then presented to respondents to replay the questions in writing. It is thus; questionnaire method is a way (means) of gathering statistical details done with the use of questionnaires given to the respondents to answer.

Field observation method

It is a method of gathering primary research data which done by a researcher looking over the phenomena. It is of two types and include; **participant** and **non participant observation**.

Scheduling method

This method of data collection is very much familiar to questionnaire. But it has little difference to questionnaire. The difference is that, schedule involves a prepared set of questions which are filled in by enumerators who are especially appointed for the purpose and of which carefully selected and trained enough to perform their job well. This method of data collection is very useful for carrying out population census. The secondary sources providing statistical data include

Literature review method

It is a systematic survey of the past documentary sources prepared by other researchers related to the study. The documentary sources include; text books, statistical abstract census report, research articles, journals, news paper, and official reports.

Other methods for data collection include; measurements, counting and the carrying out of experiments.

Strengths of statistics application in Geography

Application of statistics in geography offers the following vital significance.

- Summarizes massive information by making more simple and thus, enable the geographers to handle large sets of data.
- Statistics facilitate the process of data computation techniques possible in geography
- Statistics make easy the process of data comparison. It is so; as it is impossible to make comparison without statistics of the variables to be compared.
- Statistics application facilitates the process of drawing relationship between the geographical variables like; climate and production, population and time; rainfall and temperature etc.
- Application of statistics makes easy the process of data storage in form of numbers, tables, graphs, diagrams, and maps.
- Application of statistics makes the geographical data be clearly understood and easy for being analyzed and interpreted.
- Statistics enhance validity testing of the geographical models, theories, and concepts to the real world situations.

1.4 STATISTICAL MEASURES

Numerical values which are analyzed or examined to judge their implication (results) by taking into consideration of the statistical **measures**. It is thus; statistical measures refer to the computed numerical values used to make data analysis as related to other values in a data set provided.

Statistical measures are numerous but with regards to their nature and roles, broadly divided into the following categories.

- Measures of central tendency ·

Measures of variability

MEASURES OF CENTRAL TENDENCY.

These are the measurements which show the central values and include; arithmetic mean, mode and median.

ARITHMETIC MEAN

Arithmetic mean is an average of all values in a set of distribution. It is determined by adding up all values and divided by the sum of observation added. Arithmetic mean is used to assess the distribution value whether was high or low.

Computation of the arithmetic mean

Computation of the arithmetic mean depends up on the nature of data given whether ungrouped or grouped.

For the ungrouped data set; arithmetic mean is computed by applying the following formula

$$\text{Arithmetic mean} = \frac{\sum X}{N}$$

Where by:

$$\sum X = \text{The sum of all values in the data set given}$$

N = The total number of observation added.

Example:

Find the arithmetic mean for the following set of data.

5,7,10,12,13,14,15,7, and 2.

Solution

The arithmetic mean for the given set of data above is calculated as follow:

$$\text{Arithmetic mean} = \frac{\sum X}{N}$$

$$\sum X = X_1 + X_2 + \dots + X_R$$

$$5+7+10+12+13+14+7+2=85$$

$$\sum x = 85$$

$$N = 9$$

$$\frac{85}{9} = 9.4$$

Thus: The Arithmetic mean = 9.4

For the grouped data set; the arithmetic mean is calculated by the following application:

$$\text{Arithmetic mean} = \frac{\sum fx}{\sum f}$$

Where by; X =

Class mark f

= Frequency

Example:

Find the arithmetic mean for the following scores of marks

Class Interval	F	X	fx
91-95	0	93	0
86-90	1	88	88
81-85	6	83	498
76-80	10	78	780
71-75	15	73	1095
66-70	34	68	2312
61-65	22	63	1386
56-60	10	58	580
51-55	2	53	106

Solution:-

$$\text{Arithmetic mean} = \frac{\sum fX}{\sum f}$$

According to the given data;

$$\Sigma = 6845$$

$$\Sigma = 100$$

$$\frac{6845}{100} = 68.45$$

Thus; the arithmetic mean = 68.45

Advantages of the Arithmetic mean

- It is easy to calculate and the majority of people use to understand it
- It is used to check the values if high or low
- It can be used for further calculation. For instance; arithmetic mean is used to calculate standard deviation.

Disadvantage of the arithmetic mean

- Arithmetic mean has a big weakness of being pulled towards an outlier (extreme scores).
- It needs high mathematical knowledge to calculate arithmetic mean for the grouped data set.

MODE

Mode is a value number which occurs most frequently in a data set given .Or .

Is the most commonly attained measurement value in a data set .Or

Is the measurement value that appears most in a particular variable among a sample of subjects.

Mode helps us to know concentration of values which can stimulate scientific investigation.

Calculation of a mode

Determination of a mode is depend much up on the nature of data set whether ungrouped or grouped.

For the ungrouped data set; mode is obtained by taking the number that appears most frequently or the one that has highest frequency than the rest

Example;

Determine the mode for the following data set.

2, 4, 2, 2, 5, 6, 4

Value	Concentration
2	3
4	2
5	1
6	1

Thus; the mode for the data set given = 2

Note

Sometimes; a given data set may have more than one modes or no more at all. The one mode obtained in a set of distribution is known as unimodal or monomodal. If two modes obtained from data set; described as bimodal.

Example:

(1) 2, 5, 4, 3, 5, 6, 6, 8, 5, 6.

The modes for the data set are 5 and 6

(2) 4, 9, 8, 5, 6, 7

The given data set has no mode.

For the grouped data; mode is assessed by the following application.

$$\text{Mode} = L + \frac{t_1}{t_1 + t_2} \times (i)$$

Whereby:

- L = The lower limit of the modal class
- t_1 = The excess of the modal frequency over the frequency of the next lower class
- t_2 = the excess of the modal frequency over the frequency of the next higher class
- (i) = the class interval

Example;-

The tabled data below shows the score of marks in geography subject test form V students

Class interval	Frequency
40 – 44	7
45 – 49	8
50 – 54	11
55 – 59	10
60 – 64	4

Solution

The mode for the given data set above is calculated as follow:-

$$\text{Mode} = L + \frac{t_1}{t_1 + t_2} \times i \quad (i)$$

According to the given data

set; $L = 49.5$ $t_1 = 3$ $t_2 = 1$ $i = 5$

Then;

$$49.5 + \frac{3}{3+1} \times 5$$

$$49.5 + \frac{3}{4} \times 5$$

$$49.5 + (0.75 \times 5)$$

$$49.5 + 3.75 = 53.25$$

Thus; the mode = 53.25

Advantages of a mode

- It helps to make determination of predominance of a certain geographical feature in a place.
- It helps to know number of occurrence of the values in data set.

Disadvantages of a mode

- It needs high mathematical knowledge to calculate mode for the grouped data set
- It is unreliable measures of central tendency as a data set may have more than one modes or no mode at all.

MEDIAN

Median refers to a point value that divides the other values in a set of distribution into two equal parts after to have been arranged in ascending or descending order.

Computation of the median

The computation of the median chiefly depends on the nature of data set given if ungrouped or grouped.

For the ungrouped data set, the calculation of median should further take into account the nature of data set given whether odd or even.

If the ungrouped data set is odd; the median is just the middle value and it is obtained after the value numbers to have been arranged in ascending or descending order. E.g.

1, 2, 1, 4, 6, 5, 3

Solution

The ascending order of the values is as follow:-

1, 1, 2, 3, 4, 5, 6

Thus; the median = 3.

If the data set is even; median **is the average of the two middle values** and obtained after the value numbers to have been arranged in ascending descending order. E.g.

1,4,5,2,7,8,3,2

The ascending order for the values is as follows:-

1,2,2,3,4,5,7,8

$$\frac{3+4}{2} = \frac{7}{2} = 3.5$$

Thus; the median = 3.5

Median determination for the grouped data

For the grouped data; median is determined by applying the following formula:-

$$\text{Median} = L + \left\{ \frac{\frac{N}{2} - n_b}{n_w} \right\} * i$$

Where by:-

L = The lower limit of the median class N = Total number of observation
 n_b = the number of elements in the classes below the median class
 n_w = number of elements in the median class i = class interval

Example:-

The tabled data below: shows the score of marks in geography subject for form V students.

Class interval	Frequency
40 - 44	7
45 - 49	8
50 - 54	11
55 - 59	10
60 - 64	4

Example:-

The tabled data below; shows the score of marks in geography subject for form V students.

According to the given data

L = 49.5 N = 40 n_b = 15 n_w = 11 i = 5 n_b = the number of

elements in the classes below the median class n_w = number of

elements in the median class i = class interval

$$\frac{49.5 + (20-15) \times 5}{11}$$

$$\frac{49.5 + 5}{11}$$

$$49.5 + (0.45 \times 5)$$

$$49.5 + 2.25 = 51.75$$

Thus the median = 51.75

Advantages of median

- It helps to understand the middle value among of the numerous values in a certain data set.
- It is easy to make determination particularly for the simple data set.

Disadvantages of the median

- If the values are numerous, it becomes cumbersome to arrange in ascending or descending order to get the median
- It needs high skill to determine median for the grouped data set.

MEASURES OF VARIABILITY

These are the ones which assess the variation of values in data set. The common measures of variability include the following:-

- Range
- Standard deviation
- Variance
- Mean deviation

RANGE

Range is the difference between highest and lowest values in a given set of distribution. It is used to assess the existing variation between the highest score and lowest score.

Calculation of the range

Calculation of a range also considers the nature of a data set given whether ungrouped or grouped.

For the ungrouped data set, range is calculated by subtracting the lowest value from the highest value in a data set given.

Example:-

Determine the range for the following data set 4, 2, 3,5, 6,4, 8

Solution

The range for the data set given is computed as following:-

$\text{Range} = \text{Highest value} - \text{lowest value}$

According to the given data set:-

- Highest value = 8
- Lowest value = 2
- $8 - 2 = 6$
- Thus; The range = 6
- With the result of range; If it is high implies greater variation. If the range is small, it implies there is small variation.

For the grouped data; range is calculated by subtracting the lowest class mark from the highest subtracting the lowest lower boundary from the highest lower boundary or by subtracting the lowest higher boundary from the highest higher boundary.

Example:-

Determine the range for the following data set.

Class interval
10 – 14
15 – 19
20 – 24
25 – 29
30 – 34
35 - 39

Solution

The range for the data set given is calculated as follow:

$$\text{Range} = \text{Highest class mark} - \text{Lowest class}$$

Determination of the class mark

Class interval	Class marks
10 – 14	12
15 – 19	17
20 – 24	22
25 – 29	27
30 – 34	32
35 – 39	37

According to the computed class marks

- Highest class mark = 37
- Lowest class mark = 12
 $37 - 12 = 25$,

Thus, the range = 25

Advantages of a range

- Range gives a quick rough estimate of variability
- It is simple to calculate and the majority are much aware with it.

Disadvantages of a range

- It considers only two values of highest and lowest and thus not sensitive to the total distribution
- It is affected by the extreme values

STANDARD DEVIATION

Deviation is the difference between the value and the mean. It is computed by subtracting a the mean from the value.

$$\text{Deviation} = \bar{X} - X$$

Whereby:-

X = value given in a set of distribution

\bar{X} = average of all values

Standard deviation refers to the common difference of all values from the mean. It is the root mean square deviation from the mean. It is the measure which determines how far or scattered are the values from the mean.

Standard deviation is represented by sigma symbol of δ .

Computation of a standard deviation

Calculation of a standard deviation also depends on the nature of dataset given whether ungrouped or grouped.

For the ungrouped data; standard deviation is calculated by the following application.

$$\text{Standard deviation} = \sqrt{\frac{\sum(X - \bar{X})^2}{N}}$$

Where by:-

X = value in a set of distribution

\bar{X} = Average of all values in a set of distribution

N = The total number of observation

Example:-

Calculate the standard deviation for the following data set.

3, 2, 1, 4, 6

Solution

$$\text{Standard deviation} = \sqrt{\frac{\sum(X - \bar{X})^2}{N}}$$

Mean determination

$$\text{mean} = \frac{\sum X}{N}$$

$$\frac{3 + 2 + 1 + 4 + 6}{5} = \frac{16}{5} = 3.2$$

X	3	2	1	4	6
X- \bar{X}	-0.2	-1.2	-2.2	0.8	2.8
X-X ²	0.04	1.44	4.84	0.64	7.84

$$\sum (X - \bar{X})^2 = 11.8488$$

Then;

$$\sqrt{\frac{11.848}{5}}$$

$$\sqrt{2.37696} = 1.541$$

Hence; The SD = 1.541

For the grouped data set; standard deviation is computed by the following application:-

$$SD = \sqrt{\frac{\sum fd^2}{N}}$$

$$d^2 = (X - \bar{X})^2$$

Example:-

Calculate the SD for the following set of grouped data.

Class interval	Frequency
40 – 44	7
45 – 49	8
50 – 54	11
55 – 59	10
60 – 64	4

Procedure:

Determination of the mean

$$\text{Mean} = \frac{\sum fX}{\sum f}$$

Class interval	F	X	Fx
40 – 44	7	42	294
45 – 49	8	47	376
50 – 54	11	52	572
55 – 59	10	57	570
60 – 64	4	62	248

$$\sum fx = 2060$$

$$\frac{2060}{40} = 51.5$$

Hence; 51.5

Then:-

X	42	47	52	57	62
X – X	-9.5	-4.5	0.5	5.5	10.5
(X-X) ²	90.25	20.25	0.25	30.25	110.25
F(X – X) ²	631.75	162	2.75	302.5	441

$$\sum fd^2 = 1540$$

$$\sum fd = 40$$

$$\sqrt{\frac{1540}{40}}$$

$$\sqrt{38.5} = 6.204$$

Thus; The SD = 6.204

Note:-

The square root of SD is known as variance. Its computation is done by the following applications which also consider the nature of data set whether ungrouped or grouped.

For the ungrouped data; variance is computed by the following application:-

$$\text{Variance} = \frac{(\bar{x} - \mu)^2}{N}$$

MEAN DEVIATION

Mean deviation is the average of all deviation values. Or is the amount by which the individual values deviate from mean irrespective of its sign. It is computed by dividing the sum of all deviations irrespective of signs by the number of observation.

Calculation of mean deviation

Calculation of a mean deviation also depends on the nature of data set given whether ungrouped or grouped.

For the ungrouped data set; the mean deviation is calculated by the following application:-

$$\text{Mean deviation} = \sqrt{\frac{\sum |x - \bar{x}|}{N}}$$

Example:-

Determine the mean deviation for the following data set. 4, 7, 8, 2, 9, 6

Solution

Mean determination

$$\text{mean} = \frac{\sum x}{N}$$

$$4 + 7 + 8 + 2 + 9 + 6 = 36$$

$$\frac{36}{6} = 6$$

Hence; the mean = 6

Deviations determination

X	X - \bar{X}	D
4	4 - 6	2
7	7 - 6	1
8	8 - 6	2
2	2 - 6	4
9	9 - 6	3
6	6 - 6	0

The sum of deviations determination.

$$\sum d = d_1 + d_2 + d_3 + \text{---} \text{---} \text{---} d_n$$

$$2 + 1 + 2 + 4 + 3 + 0 = 12$$

Then; $\frac{12}{6} = 2$

Thus; the mean deviation = 2

For **the grouped data set**, mean deviation is computed by the following application:-

$$\text{Mean deviation} = \sum \frac{f |X - \bar{X}|}{N}$$

Example:-

Class interval	Frequency
40 - 44	7
45 - 49	8
50 - 54	11
55 - 59	10
60 - 64	4

Determination of the mean

$$\text{Arithmetic mean} = \frac{\sum fx}{\sum f}$$

Class interval	F	X	Fx
40 – 44	7	42	294
45 – 49	8	47	376
50 – 54	11	52	572
55 – 59	10	57	570
60 – 64	4	62	248

$$\sum fx = 2060$$

$$\sum f = 40$$

$$\frac{2060}{40} = 51.5$$

Hence; The mean = 51.5

Determination of the deviations.

$$\text{Deviation} = x - \bar{x}$$

Where by:

X = Class mark

\bar{x} = The mean

X	X - \bar{x}	D	F	Fd
42	42 – 51.5	9.5	7	66.5
47	47 – 51.5	4.5	8	36
52	52 – 51.5	0.5	8	36
57	57 – 51.5	5.5	10	55
62	62 – 51.5	10.5	4	42

The sum of (fd) determination

$$\Sigma d = fd_1 + fd_2 + fd_3 + \dots + fd_n$$

$$66.5 + 36 + 5.5 + 55 + 42 = 205$$

Then;

$$\frac{205}{40} = 5.125$$

Thus; The mean deviation = 5.125

1.5 METHODS OF PRESENTING DATA

As it has been introduced in the chapter one; the numerical data after being collected, summarized and analyzed; are presented to provide **pictorial view (visual idea)**. One of the useful ways for presenting the numerical facts is by **diagrams**. It is thus; statistical diagrams designed to illustrate values of geographical items and in turn allow quantitative analysis.

The most useful statistical diagrams for the illustration of quantitative data include the following.

- Pie chart
- Proportional semi divided circles
- Divided rectangle
- Proportional circles
- Scatter diagram
- Wind rose
- Polar chart

PIE CHART

Pie chart is also known as **divided circle** or **pie graph**. It is a method of drawing a circle of any convenient size divided proportionally into a number of segments to show the values of items in percentages. The number of segments the circle is divided into depends on the number of items whose have to appear in the circle. The proportional size of segments is determined by the degree values of the percentages.

Construction of pie chart

Consider the given data below for world production of cocoa by countries in 1968.

Country	Production
Brazil	1,936,297

Ghana	4,042,988
Nigeria	2,308,066
Ecuador	805,499
Cameroun	1,270,211
Ivory coast	1,796,883
Others	33,304,430

Procedures:-

(a) Total values determination.

$$\sum x = x_1 + x_2 + x_3 + \dots + x_n$$

$$1,936,297 + 4,042,988 + 2,308,066 + 805,499 + 1,270,211 + 1,796,883 + 33,304,430 = 15,490,374$$

(b) Percentage values determination

$$\%X = \frac{X}{\sum X} * 100$$

$$\text{Brazil: } \frac{1,936,297}{15,490,374} \times 100\% = 12.5\%$$

$$\text{Ghana: } \frac{4,042,983}{15,490,374} \times 100\% = 26.1\%$$

$$\text{Niger: } \frac{2,308,066}{15,490,374} \times 100\% = 14.9\%$$

$$\text{Ecuador: } \frac{805,499}{15,490,374} \times 100\% = 8.2\%$$

$$\text{Cameroun: } \frac{1,270,211}{15,490,374} \times 100\% = 11.6\%$$

$$\text{Ivory Cost: } \frac{1,796,883}{15,490,374} \times 100\% = 11.6\%$$

$$\text{Nigeria: } \frac{2,308,066}{15,490,374} \times 100\% = 14.9\%$$

$$\text{Others: } \frac{33,304,430}{15,490,374} \times 100\% = 21.5\%$$

(c) Degrees of the percentage values determination:

$$\text{Degree values} = \frac{\%}{100} \times 360^\circ$$

$$\text{Brazil : } \frac{12.5}{100} \times 360^\circ = 45^\circ$$

$$\text{Ghana : } \frac{26.1}{100} \times 360^\circ = 93.6^\circ$$

$$\text{Nigeria: } \frac{14.9}{100} \times 360^\circ = 53.6^\circ$$

$$\text{Ecuador: } \frac{5.2}{100} \times 360^\circ = 18.7^\circ$$

$$\text{Cameroun : } \frac{8.2}{100} \times 360^\circ = 29.5^\circ$$

$$\text{Ivory cost : } \frac{11.6}{100} \times 360^\circ = 41.8^\circ$$

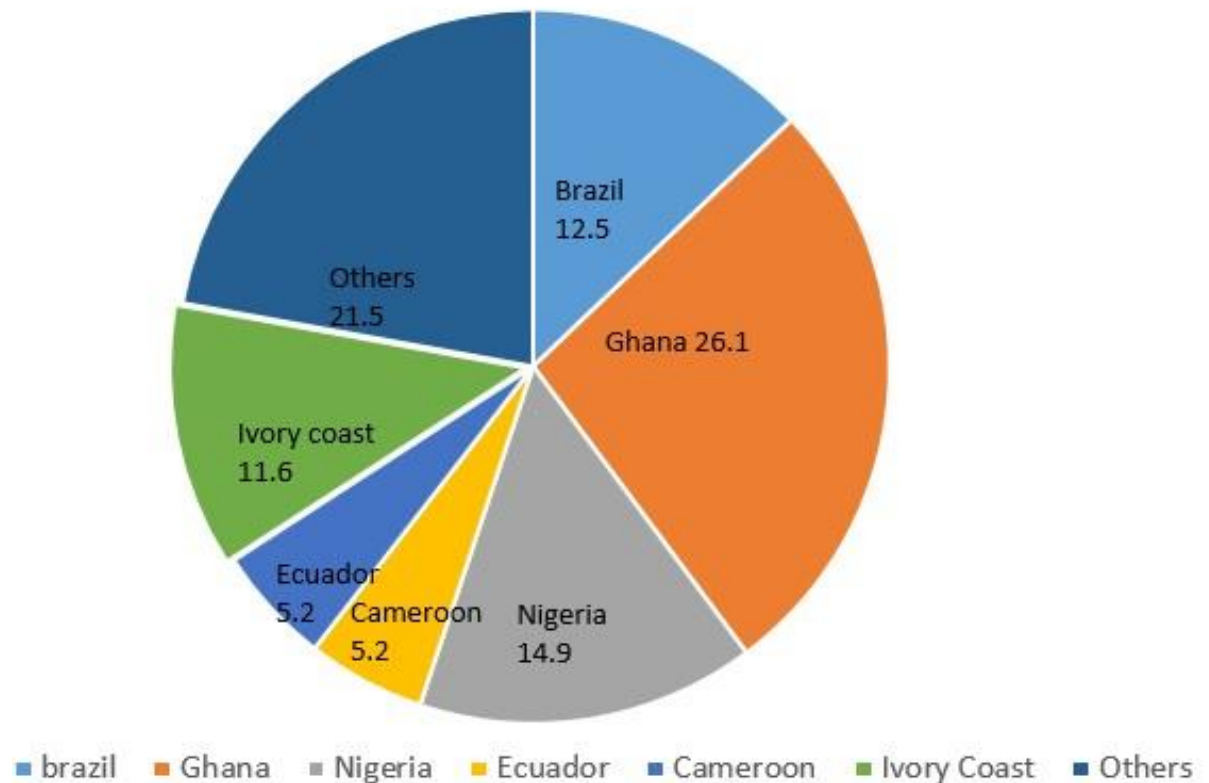
$$\text{Others : } \frac{21.5}{100} \times 360^\circ = 77.4^\circ$$

Country	Production	X%	X ⁰
Brazil	1,936,297	12.5%	4 ⁰
Ghana	4,042,988	26.1%	93.6 ⁰
Nigeria	2,308,066	14.9%	53.6 ⁰
Ecuador	805,499	5.2%	18.7 ⁰
Cameroun	1,270,211	5.2%	29.5 ⁰
Ivory coast	1,796,883	11.6%	41.8 ⁰
Others	33,304,430	21.5%	77.4 ⁰

- (d) The circle of any convenient size should be drawn. It should be divided into proportional segments with respect to the computed degree values. Too small circle is not required.

It is thus; the pie chart for the data appears as follow.

WORLD PRODUCTION OF COCOA BY COUNTRIES – 1968



Strengths of the pie chart

- The method is more pleasing to eye and it is one among the most popular methods in statistics for data representation.
- The values given by the method are more simplified as appear in percentages.
- It allows the easy making of quantitative analysis.

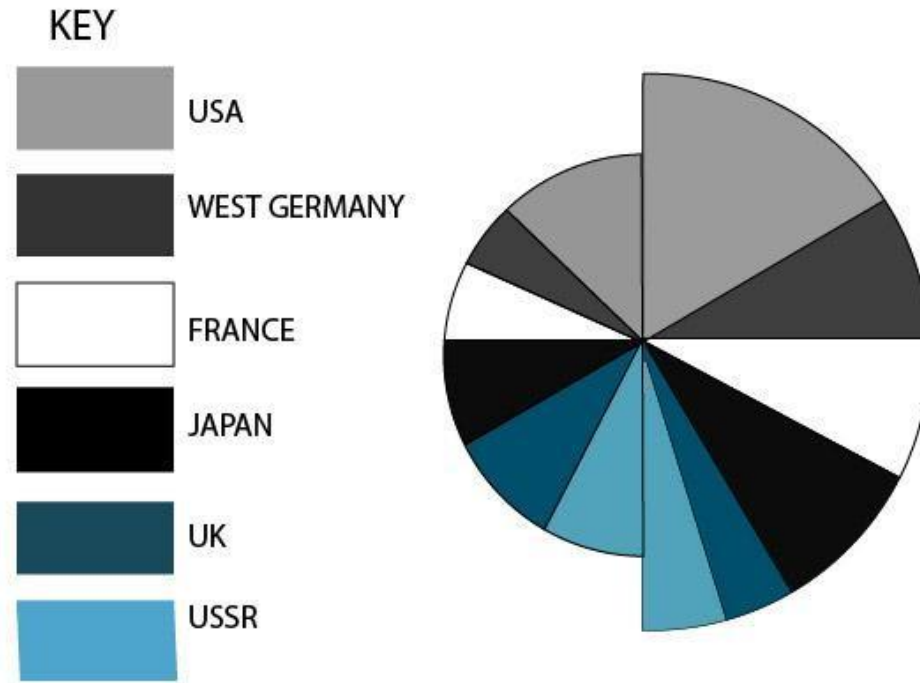
Setbacks of the pie chart

- It does not give the absolute values of items represented
- It consumes much of time to prepare. Hence it is tedious enough.
- It needs high skill to prepare it
- A problem may arise in selecting the varied shade of textures.

PROPORTIONAL SEMI DIVIDED CIRCLE

This pictorial method, involves the drawing of two semi circles linked to one another and each is proportional to the total quantity represented. Each semi circle is proportionally divided into segments and the number of segments the semi circle is made to have, depends on the number of items. In making the segments in the semi circle, 180^0 is used as the total degree for each semi cycle.

The method is very useful in making comparison of items for two major cases like dates or places.



Construction of the proportional semi divided circles

Consider the following tabled data showing motor vehicles production for passengers and commercial in the industrialized nations.

	<i>W. total</i>	<i>USA</i>	<i>W. Germany</i>	<i>Japan</i>	<i>U.K.</i>	<i>France</i>	
Comm	6530	1896	241	2052	409	242	750
Passe	21720	8222	2862	2055	1816	1832	280

Procedure:-

(a) Find the total values for each variable

$$\text{Commercial: } 1896 + 241 + 2052 + 409 + 242 + 750 + 940 = 6530$$

$$\text{Passengers: } 8222 + 2862 + 2055 + 1816 + 1832 + 280 + 4653 = 21720$$

(b) Angle determination of the segments

Commercial:-

$$\text{USA} : \frac{1896}{6530} \times 180^\circ = 52.3^\circ$$

$$\text{West germany} : \frac{241}{6530} \times 180^\circ = 6.6^\circ$$

$$\text{Japan} : \frac{1896}{6530} \times 180^\circ = 56.6^\circ$$

$$\text{UK} : \frac{409}{6530} \times 180^\circ = 11.3^\circ$$

$$\text{France} : \frac{242}{6530} \times 180^\circ = 6.7^\circ$$

$$\text{USSR} : \frac{750}{6530} \times 180^\circ = 20.7^\circ$$

$$\text{Others} : \frac{940}{6530} \times 180^\circ = 6.6^\circ$$

(c) Passengers.

$$\text{USA} : \frac{8222}{21720} \times 180^\circ = 63.1^\circ$$

$$\text{West germany} : \frac{2862}{21720} \times 180^\circ = 23.7^\circ$$

$$\text{Japan} : \frac{2055}{21720} \times 180^\circ = 17^\circ$$

$$\text{UK} : \frac{1816}{21720} \times 180^\circ = 15^\circ$$

$$\text{France} : \frac{1832}{21720} \times 180^\circ = 63.1^\circ$$

$$\text{USSR} : \frac{280}{21720} \times 180^\circ = 2.3^\circ$$

$$\text{Others} : \frac{4653}{21720} \times 180^\circ = 38.6^\circ$$

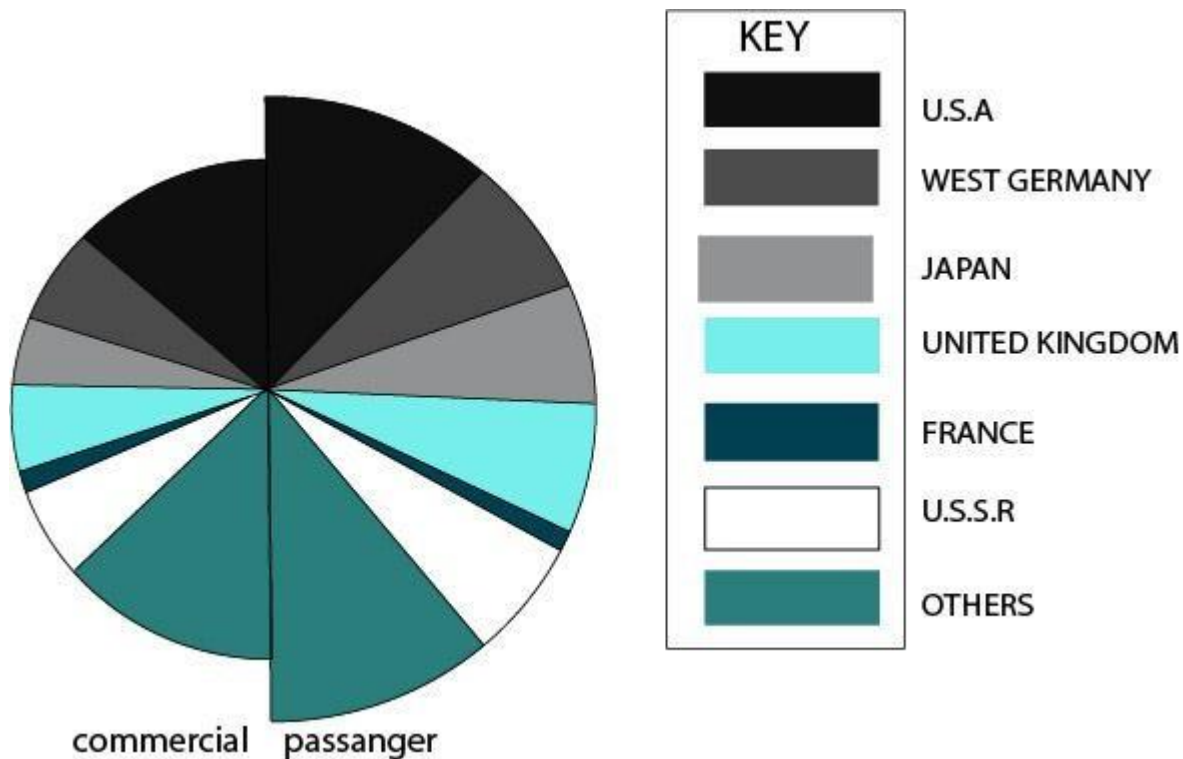
- (d) Estimation of the diameters of the two proportional semi circles. It is much up on specific scale. The scale is developed by proposing specific value to be represented by 1cm. Le say 1cm should represent 20000 motor vehicles.

$$\text{Passenger} : \frac{21720}{2000} = 10.86 \text{ cm}$$

$$\text{Commercial} : \frac{6530}{2000} = 3.2 \text{ cm}$$

Thus the proportional semi divided circle for the data given appears as follow:-

WORLD PRODUCTION OF MOTOR VEHICLES BY COUNTRIES (000)



Diameter scale:-

1cm represents 2000 motor vehicles.

Merits of the semi divided proportional circles.

- It is useful technique for showing comparison of item values for two major cases · It provides visual idea
- It allows the making of quantitative analysis

Setbacks of the proportional semi divided circles.

- It needs high skill to extract actual values from the diagram
- It consumes much time to prepare
- It needs high skill to be prepared
- It encounters a problem shade textures selection

DIVIDED RECTANGLE

It is one among the most useful and versatile method of statistical presentation of data. However it is not frequently used. By this method, the total quantity is presented by a rectangle which is then sub divided to represent the constituent parts.

Depending on the function, the divided rectangle is of two types including:-

- Simple divided rectangle
- Compound divided rectangle

SIMPLE DIVIDED RECTANGLE

It is a rectangle drawn to have a length proportional to the total quality represented, then divided into proportional segments to show the values of the cases.

Construction of the simple divided rectangle

Consider the following data of coffee production in Tanzania in '000' tons in 1980.

REGION	PRODUCTION
Arusha	9
Kilimanjaro	10
Bukoba	18
Ruvuma	11
Mbeya	7
Tanga	5

- (a) Determine the scale value to be used in drawing the rectangle.

$$\text{Scale} = \frac{\text{Cumulative value}}{\text{Rectangle length}}$$

$$\frac{60}{10} = 6$$

Hence; 1 cm represents 6 tons.

- (b) Determine the length of the values along the rectangle

$$\text{Proportional Length} = \frac{\text{Value}}{\text{Scale value}}$$

Arusha : $\frac{9}{6} = 1.5 \text{ cm}$

Kilimanjaro : $\frac{10}{6} = 1.7 \text{ cm}$

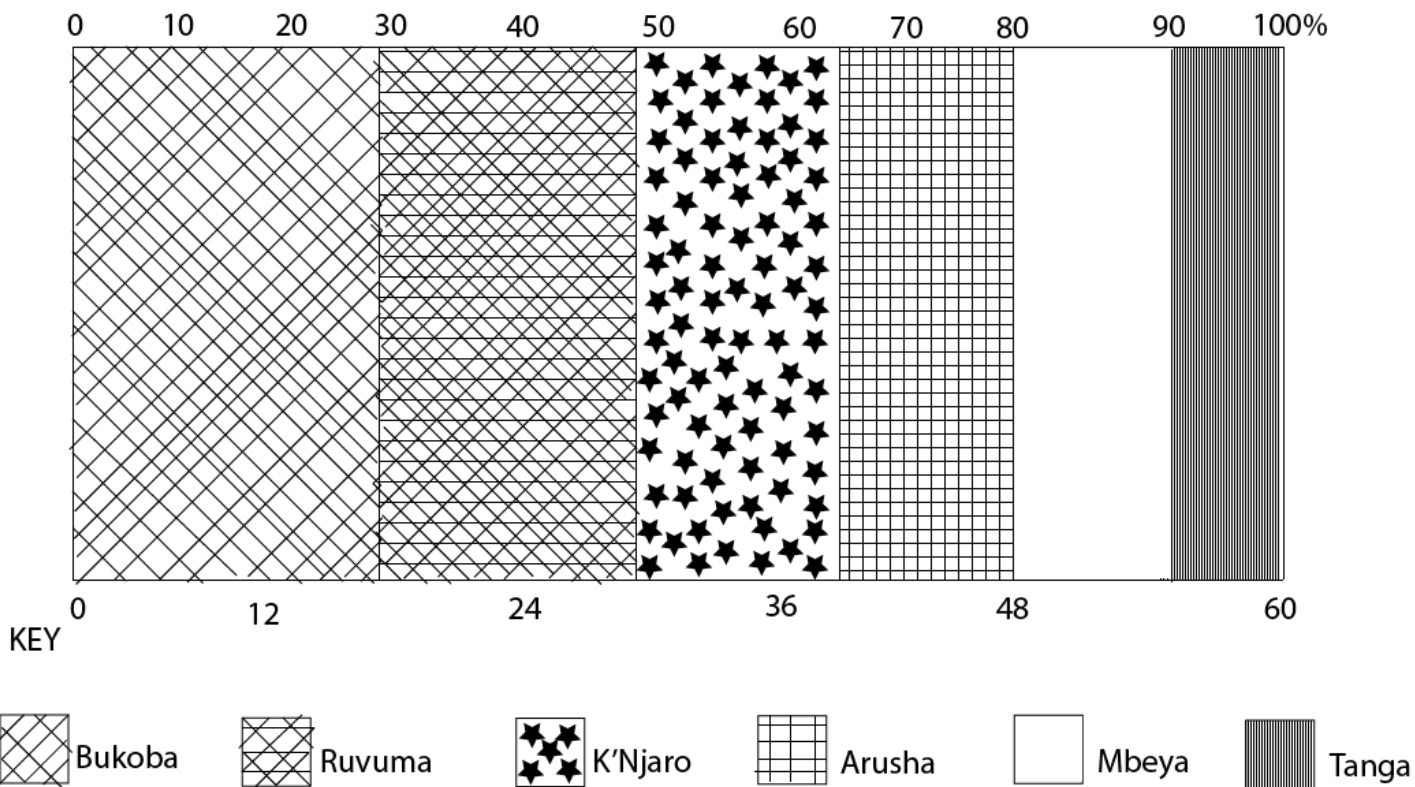
Bukoba : $\frac{18}{6} = 3 \text{ cm}$

Ruvuma : $\frac{9}{6} = 1.5 \text{ cm}$

Mbeya : $\frac{7}{6} = 1.2 \text{ cm}$

Tanga : $\frac{5}{6} = 0.8 \text{ cm}$

Thus; the simple divided rectangle for the given data appears as follow:-



Simple divided rectangle coffee production in '000' tons from 1980 to 1985

COMPOUND DIVIDED RECTANGLE

By the compound divided rectangle, each proportional strip in the rectangle is also proportionally divided to show further information of the cases represented. This is drawn with two scales. One

scale is for horizontal dimension, and it is designated as the horizontal scale; the other is for the vertical dimension and is designated as vertical scale. It is much better for the two scales graduated in separate values. The horizontal scale is absolute values and the vertical scale be in percentage.

Example:-

Consider the given data below showing land use partners for the six village:-

COUNTRIES	SIZE LANDUSE OF TOTAL AREA '000' KM ²			
	Westland	Pasture	Arable	forestry
Ruvu Darajani	166.5	27	31.5	225
Vigwaza	94.4	-	40.4	202.2
Buyuni	226.8	-	32.4	64.8
Kidogozero	7.7	5.2	21.5	8.6
Visezi	8.5	13.3	6.8	5.4
Kitonga	8.8	8.2	10	7

Procedure:-

(a) Cumulative values determination.

Ruvu Darajani:- $166.5 + 27 + 31.5 + 225 = 450$

Vigwaza:- $94.4 + 40.4 + 202.2 = 337$

Buyuni:- $226.8 + 32.4 + 64.8 = 324$

Kidogezero:- $7.7 + 5.2 + 21.5 + 8.6 = 43$

Visezi:- $8.5 + 13.3 + 6.8 + 5.4 = 34$

Kitonga:- $8.8 + 8.2 + 10 + 7 = 34$

(b) The percentage values determination

· **Ruvu Darajani:-**

$$\text{Waste land : } \frac{166.5}{450} \times 100\% = 37\%$$

$$\text{Pasture : } \frac{27}{450} \times 100\% = 6\%$$

$$\text{Arable : } \frac{31.5}{450} \times 100\% = 7\%$$

$$\text{Forestry : } \frac{225}{450} \times 100\% = 50\%$$

· **Vigwaza:-**

$$\text{Waste land : } \frac{94.4}{337} \times 100\% = 12\%$$

$$\text{Arable : } \frac{40.4}{337} \times 100\% = 28\%$$

$$\text{Forestry : } \frac{202.2}{337} \times 100\% = 60\%$$

· **Buyuni:-**

$$\text{Waste land : } \frac{226.8}{324} \times 100\% = 70\%$$

$$\text{Arable : } \frac{32.4}{324} \times 100\% = 10\%$$

$$\text{Forestry : } \frac{202.2}{337} \times 100\% = 60\%$$

· **Kidogezero:-**

$$\text{Waste land : } \frac{7.7}{43} \times 100\% = 18\%$$

$$\text{Pasture : } \frac{40.4}{337} \times 100\% = 12\%$$

$$\text{Arable : } \frac{21.5}{43} \times 100\% = 50\%$$

$$\text{Forestry : } \frac{8.8}{43} \times 100\% = 20\%$$

· **Kitonga:-**

$$\text{Waste land : } \frac{8.8}{34} \times 100\% = 26\%$$

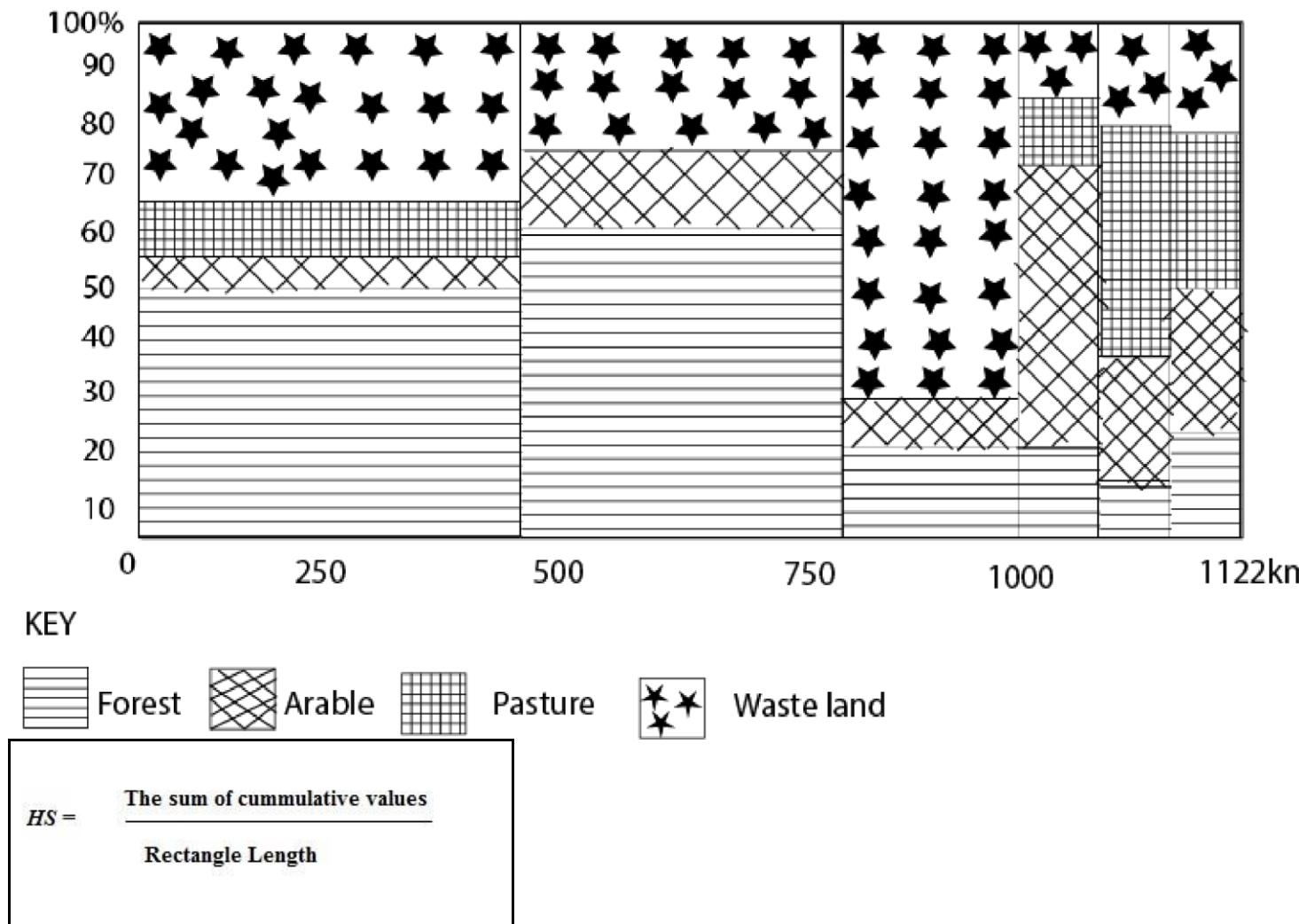
$$\text{Pasture : } \frac{8.2}{34} \times 100\% = 24\%$$

$$\text{Arable : } \frac{10}{34} \times 100\% = 29\%$$

$$\text{Forestry : } \frac{7}{34} \times 100\% = 21\%$$

(c) Scale determination

(e) Horizontal scale



$$\frac{1222}{10 \text{ cm}} = 122 - 125\%$$

Hence; Horizontal scale: 1 cm represents 125%

(f) Vertical scale

$$\frac{100\%}{10 \text{ cm}} = 10\%$$

Merits of the divided rectangle

- It is useful method for showing cumulative values

- It is more illustrative as it provides visual idea to the users in statistics
- It allows the easy making of quantitative analysis
- The data represented by compound divided graph can also be represented by percentage bar graph.

Set backs of the divided rectangle

- It is not much pleasing to people
- It consumes much time to prepare especially the compound divided rectangle
- It needs high skill to prepare the compound divided rectangle
- It needs high skill to prepare the compound divided rectangle
- It is much less used for statistical data representation
- A problem can be encountered in selecting the varied textures provided items are numerous.

PROPORTIONAL CIRCLES

It is diagram with circles whose size proportional to the quantity represented. The area size of a circle is calculated by the following application:-

$$\pi r^2 \text{ i.e. } \frac{22}{7} \times \text{radius}^2$$

But in our case; π is ignored. The radius varies with the quantity to be represented. Hence; proportional circles are drawn with radii proportional to the square root of the quantity represented.

Construction of proportional circles

Consider the given data below:-

Hydroelectricity production for some stations in country X.

HEP Station	Production in MW
A	100
B	144

C	255
D	400
E	625

Procedure:

- (a) The values should be arranged in ascending or descending order. i.e. 100,144, 255, 400, 625.
- (b) Find the square roots of the values.

$$\sqrt{100} = 10$$

$$\sqrt{255} = 15$$

$$\sqrt{400} = 20$$

$$\sqrt{625} = 25$$

- (c) Estimate the radius value scale to be used for all proportional circles. In the estimation, propose the highest radius to be used. Then the highest square root should be divided by the proposed highest radius.

$$\text{Value Scale} = \frac{\text{Highest square root}}{\text{Max Radius Length}}$$

$$\frac{25}{5} = 5$$

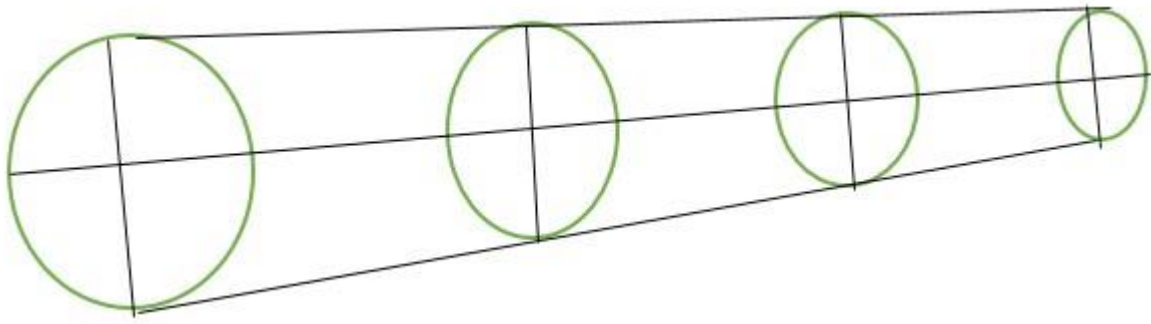
Thus; 1 cm to 5 square root.

- (d) The proportional circles should be drawn accordingly.

In *drawing* the proportional circles; the following procedure should be followed.

- The circles are drawn proportionally to the quantity represented depending on the scale that has been decided.
- The two perpendicular lines should be drawn to follow the arrangement of the circles.
- The central line should be drawn through all circles.

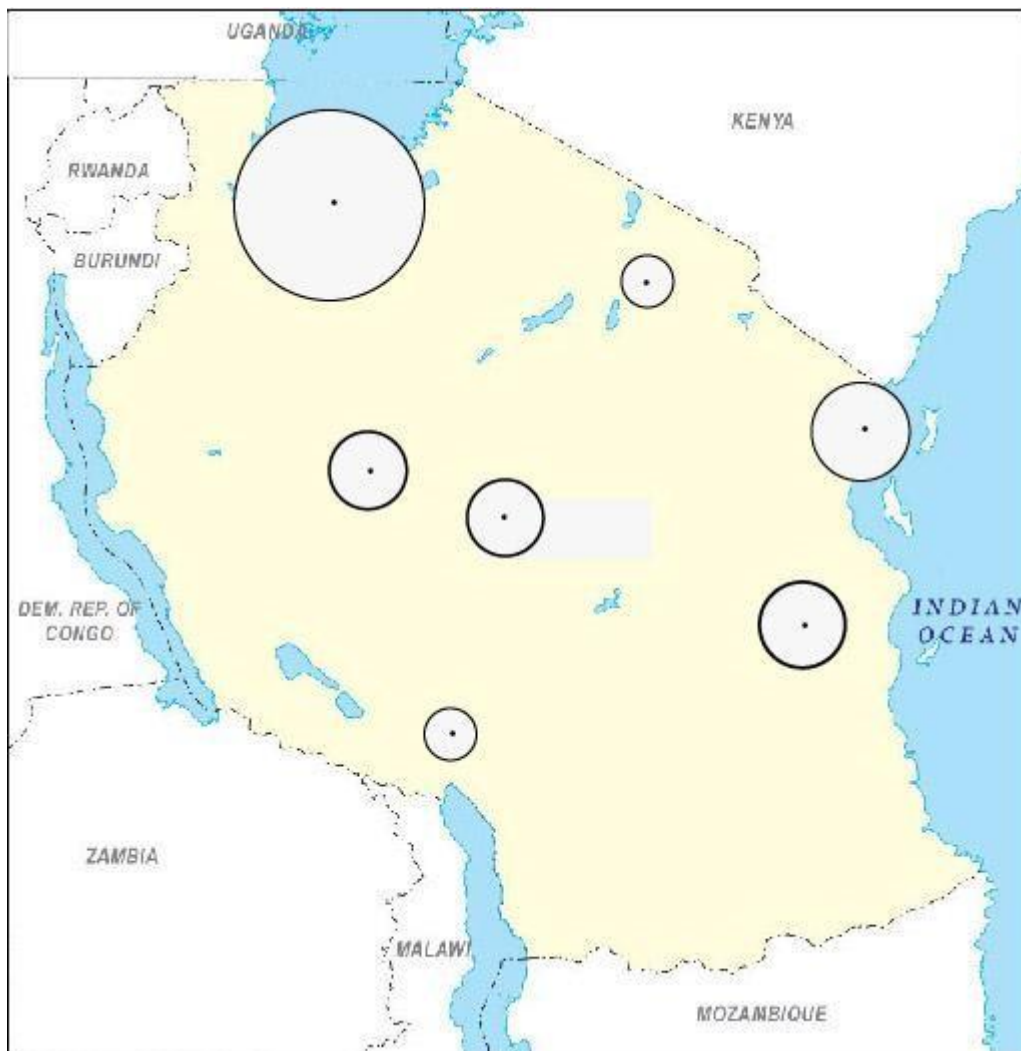
PROPORTIONAL CIRCLES SHOWING HEP PRODUCTION FOR THE STATIONS



Note

The proportional circles can be drawn on a map. This is done under the recommendation of showing values of places which appear on the map. The proportional circles on the map, sometimes may overlap. This is not a problem. But if it is possible, the best should be tried to minimize the size of the circles. One of the ways is to minimize the scale size.

Consider the map with proportional circles on the next page.



Advantages of proportional circles

- It is a good method of comparing absolute values
- The proportional circles give good visual impression

Disadvantages of the proportional circles

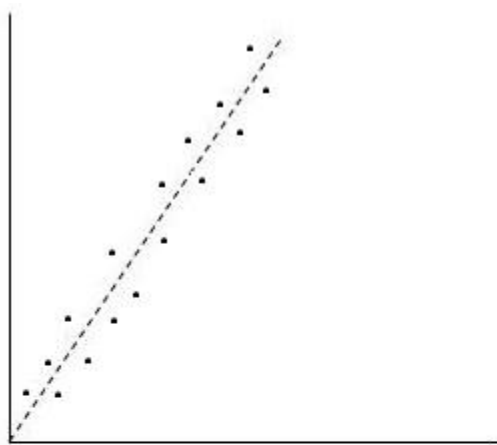
- It is much tedious in construction
- It becomes difficult to determine the exact values from the circles.

SCATTER DIAGRAM

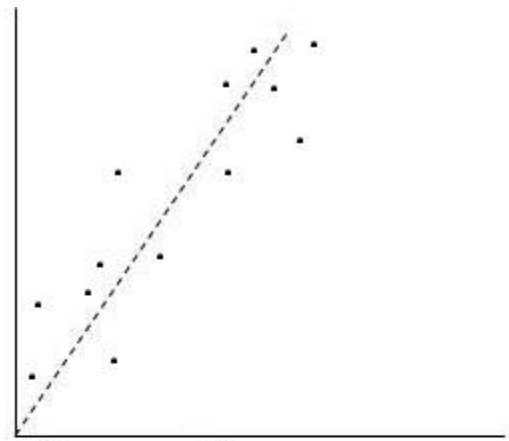
This method is also known as scatter graph. It is a statistical diagram designed to show correlation between two types of data. The diagram is made to have two lines axis. The vertical axis is used to show the values for the dependent variable; while the horizontal axis is used to show the values for independent variable.

On the diagram; a straight line is drawn to follow the distribution of dots.

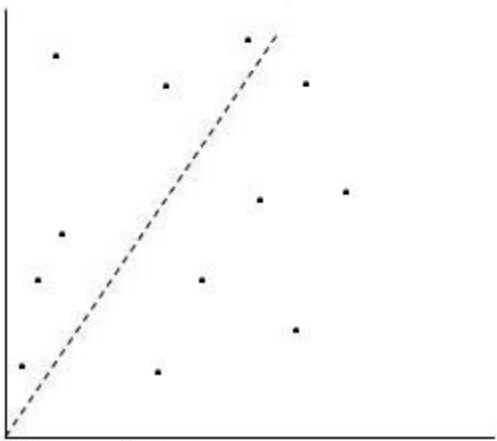
- If the plotted dots appear closer to straight line, indicates greater correlation
- If the plotted dots appear widely scattered from the line indicates low or no correlation.



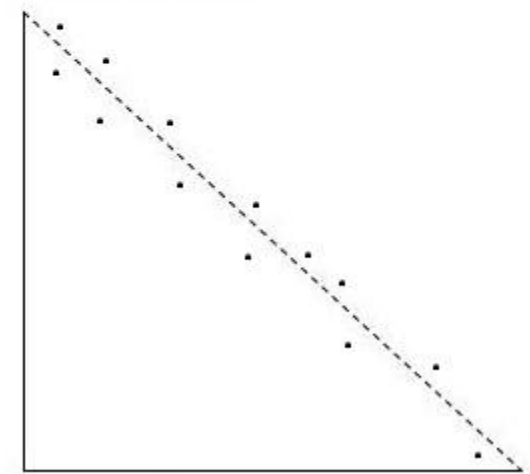
Greater correlation



Lower correlation



No correlation



Negative correlation

Construction of the scatter diagram

Consider the given data below showing the amount of rainfall at varied altitudes.

Altitude (m)	Rainfall (mm).
500	600
600	800
700	1200
800	1500
900	1700
1000	2000
1100	2400

Procedure:

(a) Identify the variables

- Dependent variable - Rainfall distribution values
- Independent variable - Altitude

(b) Estimation of both vertical and Horizontal scales

$$VS = \frac{\text{Highest Value}}{\text{Graph Space}}$$

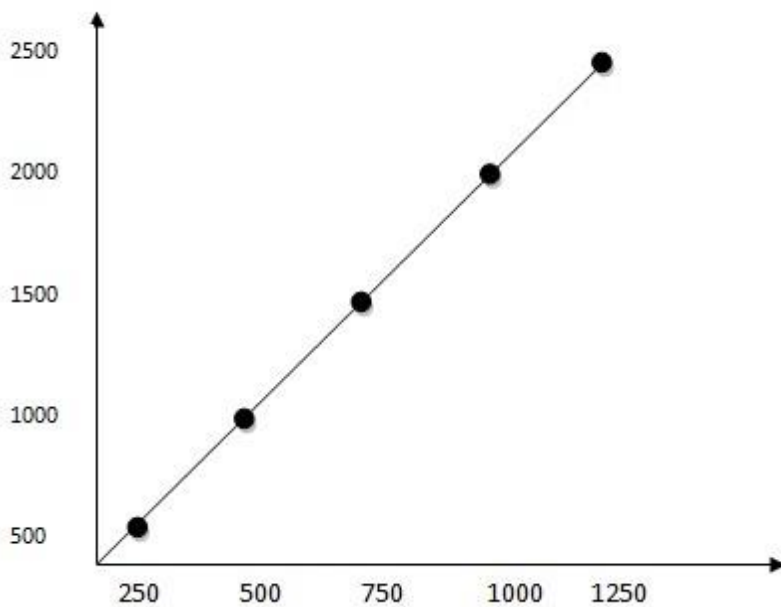
$$\frac{2400}{5 \text{ cm}} = 480 - 500\text{mm}$$

Hence; VS 1cm represents 500mm

$$HS = \frac{\text{Highest Value}}{\text{Graph Space}}$$

$$\frac{1100}{5 \text{ cm}} = 220 - 250$$

Hence; HS 1cm represents 250m





According to scatter diagram above, the plotted dots are much closer to the line, This shows greater positive correlation between rainfall and altitudes. i.e. rainfall greatly influenced by altitude.

WIND ROSE

It is a statistical diagram designed to show the number values of wind blow frequencies per varied direction and speed in a given month as it was recorded at a certain weather station.

Wind rose is of two types including simple and compound wind roses.

SIMPLE WIND ROSE

Simple wind rose only shows number of wind blow frequencies per directions. It is made to have octagon sides or a circle of any convenient size. If octagon used; on each side, a rectangle of equal or varied length to others is drawn to represent the directions from which winds were blowing. If rectangles are made to have equal length, in each, small lines established to represent the number of wind blow frequencies. If are made of not equal length, each whose length is made proportional to the number of wind blow frequencies. The number of days which didn't experience wind blow (calm days) written in a circle inside the octagon.

Example:

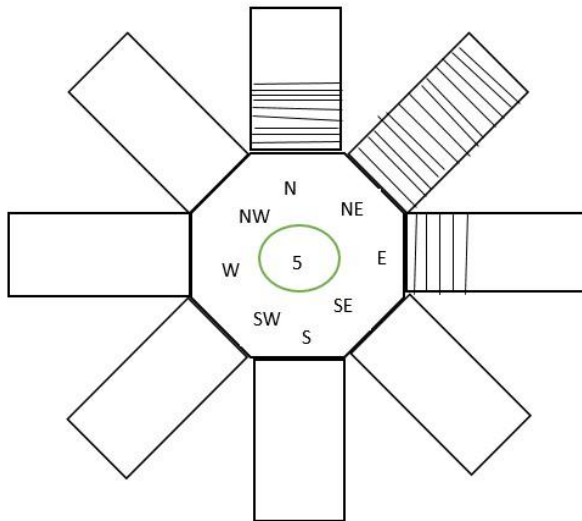
Construct the simple wind rose to represent the following data.

Wind blow frequencies at X weather station for the month of June.

DIRECTION	N	NE	E	SE	S	SW	W	NW
-----------	---	----	---	----	---	----	---	----

WIND Fq	5	4	4	1	3	4	3	
---------	---	---	---	---	---	---	---	--

WIND ROSE FOR X WEATHER STATION



COMPOUND WIND ROSE

Compound wind rose is employed to show the average wind blow frequencies per varied direction and speed commonly in percentage of a given month for station weather station.

Example:

Construct the compound wind rose to present the following data.

Wind blow frequencies at X weather station for the month of June in percentages.

Wind speed/Direction	N	NE	E	SE	S	SW	W	NW
Less than 4kph	2	2	3	3	4	2	5	4
4 – 12 kph	3	4	2	5	2	3	4	2
13 – 22 kph	2	2	1	1	3	2	2	3
Total	11	10	9	11	11	10	15	11

Calm days = 18%

The compound wind rose for the given data is constructed as follow.

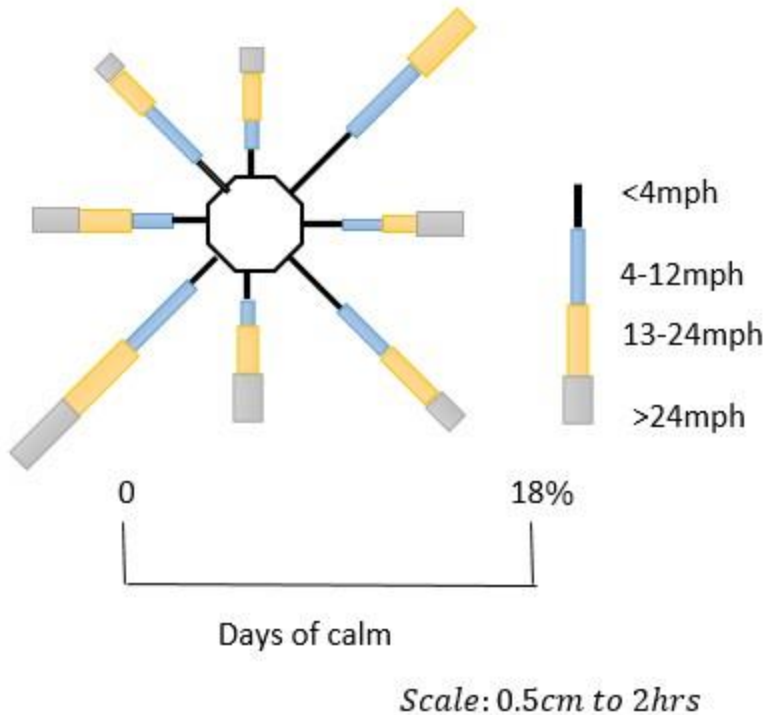
- Scale value determination

$$\text{Value Scale} = \frac{\text{Highest commulative frequency}}{\text{Graph space}}$$

$$\frac{15}{3} = 5$$

Hence; 1cm represents 3% frequency

Thus; the wind rose appears as follow:-



Advantages of wind rose:

- It gives a visual impression of wind frequencies
- It is relatively easy to construct and takes a short time provided a scale is well assessed
- It is easy to understand information represented.

Disadvantages of a wind rose

- Numerical values not easily extracted as it needs measuring and calculating using the given scale.
- One cannot know the exact time or day when wind blew from a particular, direction since the wind rose is a summary of the conditions over a period of time.
- The pattern of wind blow over a given period cannot easily be seen from the diagram.

POLAR CHART

The graph is also known as circular graph or clock graph. It is a graph in circular form designed to have bars and circular line to show two attributes whose values appear in varied unit. It is basically employed to illustrate the amount of temperature and rainfall together in a year. However polar chart can also be used in other cases of distribution recorded in a year.

For the case of showing climatic records, polar chart employ the use of both bars and line to illustrate rainfall and temperature values respectively.

The circle is divided into twelve equi angular radii.

Construction of the circular graph

The following tabled data show the climatic condition for certain weather station in Jerusalem.

Month	J	F	M	A	M	J	J	A	S	O	N	D
Temp °C	8	9.1	12.2	17	21	22	23	24	23	21	17	12
Rain mm	150	160	70	30	18	00	00	00	00	22	80	90

Steps

Estimation of the value scales to be used.

$$\text{Rain value scale} = \frac{\text{Highest value}}{\text{Graph pace}}$$

$$\frac{182mm}{5cm} = 36.4 - 40mm$$

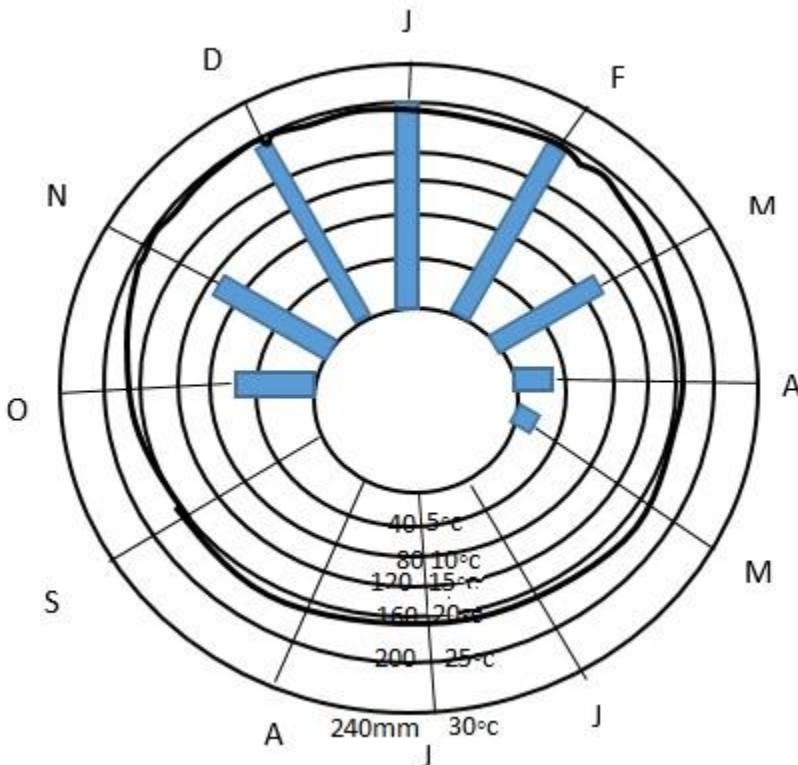
Thus, the value scale for rain fall is 1cm to 40 mm

$$\text{Temperature value scale} = \frac{\text{Highest value}}{\text{Graph pace}}$$

$$\frac{24^{\circ}C}{5cm} = 4.8 - 5^{\circ}$$

Hence; the temperature vertical scale; 1cm represents 5⁰c

The polar chart has to be drawn as follow:-



Strength as of the circular graph

- It is useful graphical method for showing the distribution values of climate
- It is more illustrative, as it provides visual idea to the users in statistics
- It allows the easy making of quantitative analysis

Setbacks of the circular graph

- It needs high skill to make quantitative analysis from the Graph
- It is time consuming graphical method in construction
- Needs high skill to construct the graph

STATISTICAL GRAPHS

These are the graphs designed to illustrate values of geographical items by means of lines or bars and in turn allow quantitative analysis.

The most useful statistical graphs for the illustration of values include the following.

- Line graphs
- Bar graphs

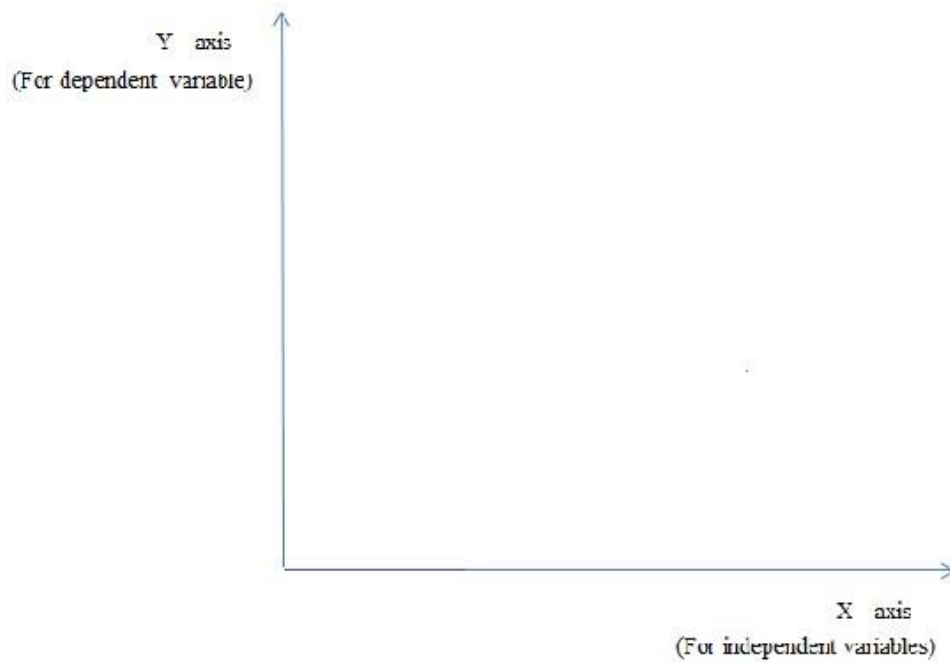
- Combined bars and line graph

LINE GRAPHS

These are the graphs which use line (s) to illustrate the values of items to give quantitative analysis.

Any line graph has two axes of the following:-

- X – axis; This is also known as **the base or horizontal axis**. It is used principally to show the value of independent variable like **date** or **places**.
- Y – axis: This is also known as **the vertical axis**. It is used show the values for the dependent variable of like output of crops, minerals etc.



TYPES OF LINE GRAPHS

Linear graphs are extremely varied. They are differently designed to meet varied functions (roles). With respect to this consideration, linear graphs recognized to be of the following forms:

- Simple line graph
- Cumulative line graph
- Divergent line graph
- Group line graph

- Compound line graph

Simple line graph

It is a form of line graph, designed to have one line to illustrate the values of one item in relation to dependent and independent variables. i.e. It is designed to show the values of one item per varied date or places.

CONSTRUCTION OF THE SIMPLE LINE GRAPH

Consider the given hypothetical data below showing maize production for country X in 0,000 metric tons (1990 – 1995).

YEAR	PRODUCTION
1990	100
1991	250
1992	300
1993	150
1994	500
1995	400

Procedure

- (a) Variables identification

Dependent variable production values

Independent variable Date (Years).

Y – axis production values

X – axis Years

- (b) Vertical and horizontal scales estimation

$$\text{Vertical scale} = \frac{\text{Highest value}}{\text{Graph space}}$$

$$\frac{500}{10 \text{ cm}} = 1 \text{ cm to } 500,000 \text{ tons}$$

Hence; VS is 1 cm to 50000 tons.

Horizontal scale is up on decision

Hence; 1cm represents 1 year

MAIZE PRODUCTION FOR COUNTRY X IN (0,000) Metric tons



Scales:

VS: 1 cm to 50 tons

HS: 1cm to 1 year

Source:

Hypothetical data

Strengths of the simple line graph

- It is much easier to prepare as it involves no complicated mathematical works, and also a single line establishes the graph.
- From the graph, the absolute values are extracted
- It is comparatively easier to read and interpret the values
- It has perfect replacement by simple bar graph

Setbacks of the simple line graph

- It is a limited graphical method as only suited to represent the value for one item.
- Sometimes it becomes difficult to assess the vertical scale if the variation between the highest and lowest values appear wider enough.

Cumulative line graph

It is a form of line graph designed to show the accumulated total values at various dates or possibly places for a single item. This graphical method has no alternative graphical bar method as it can be compared to other linear graphical methods.

Construction of the cumulative line graph

Consider the given hypothetical data below showing maize production for country X.

YEAR	PRODUCTION
1990	50
1991	40

1992	90
1993	100
1994	90
1995	130

Procedure

(a) Variables identification

- Dependent variable production values
- Independent variable ---- Date (Years)

Y - axis Production values

X – axisYears

(b) Vertical and horizontal scales estimation

$$\text{Vertical scale} = \frac{\text{Highest value}}{\text{Graph space}}$$

(c) Determination of the cumulative values.

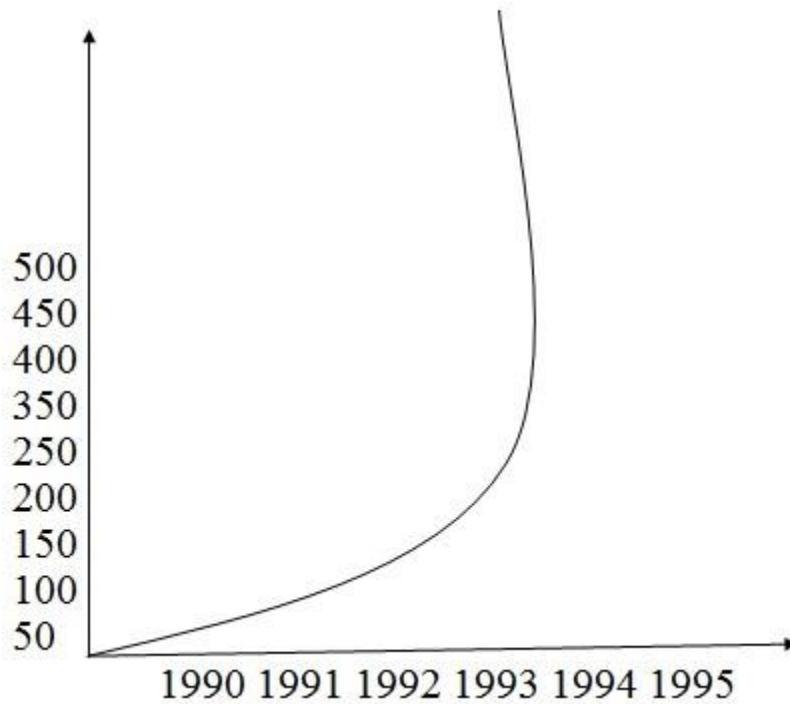
YEAR	PRODUCTION	CUM VALUES
1990	50	50
1991	40	90
1992	90	180
1993	100	280
1994	90	370
1995	130	500

$$\frac{500}{10} = 50$$

Hence: VS; 1cm represents 50 tons

Thus; the cumulative line graph appears as follow.

Cumulative line graph: Maize production for country X.



SCALE:-

VS..... 1cm represents 50 tons

HS 1cm represents 1 year

Source Hypothetical data.

Merits of the cumulative line graph

- The graphical method shows cumulative values
- From the graph the values can be revealed and quantitatively analyzed

Setbacks of the cumulative line graph

· The graphical method is not suited to show cumulative values for more than one item, it is thus; the graphical method limited for showing the values of a single item. · It needs high skill to reveal the actual values of the item represented · It has no alternative graphical bar method.

Divergent line graph

It is a form of line graph designed to illustrate the increase and decrease of the distribution values in relation to the mean. The graph is designed to have **upper and lower sections** showing **positive and negative values** respectively. The two portions are separated by the steady line graduated with zero value along the vertical line. The steady line also shows the average of all values.

Construction of the divergent line graph

Consider the following tabled data which show export values of coffee for country X in millions of dollars.

YEAR	EXPORT VALUES (000,000 dollars)
1952	345
1953	256.5
1954	283
1955	500
1956	335
1957	330.5

(a) Variables identification

- Dependent variable Export values
- Independent variable ---- Date (Years)

Y - axis Export values

X – axisYears

(b) Computation of the arithmetic mean

$$\text{Arithmetic mean} = \frac{\sum x}{N}$$

$$\sum x = x_1 + x_2 + x_3 + \dots + x_n$$

$$\cdot \quad 345 + 256 + 283 + 300 + 335 + 330.5 = 1850$$

Then;

$$\frac{1850}{6} = 308$$

Computation of the deviation values

$$\text{Deviation} = X - \bar{X}$$

$$1952 \quad 345 - 308 = 37$$

$$1953 \quad 256.5 - 308 = 52.5$$

$$1954 \quad 283 - 308 = -25$$

$$1955 \quad 300 - 308 = -8$$

$$1956 \quad 335 - 308 = 27$$

$$1957 \quad 330.5 - 308 = 22.5$$

(c) Estimation of the vertical scale.

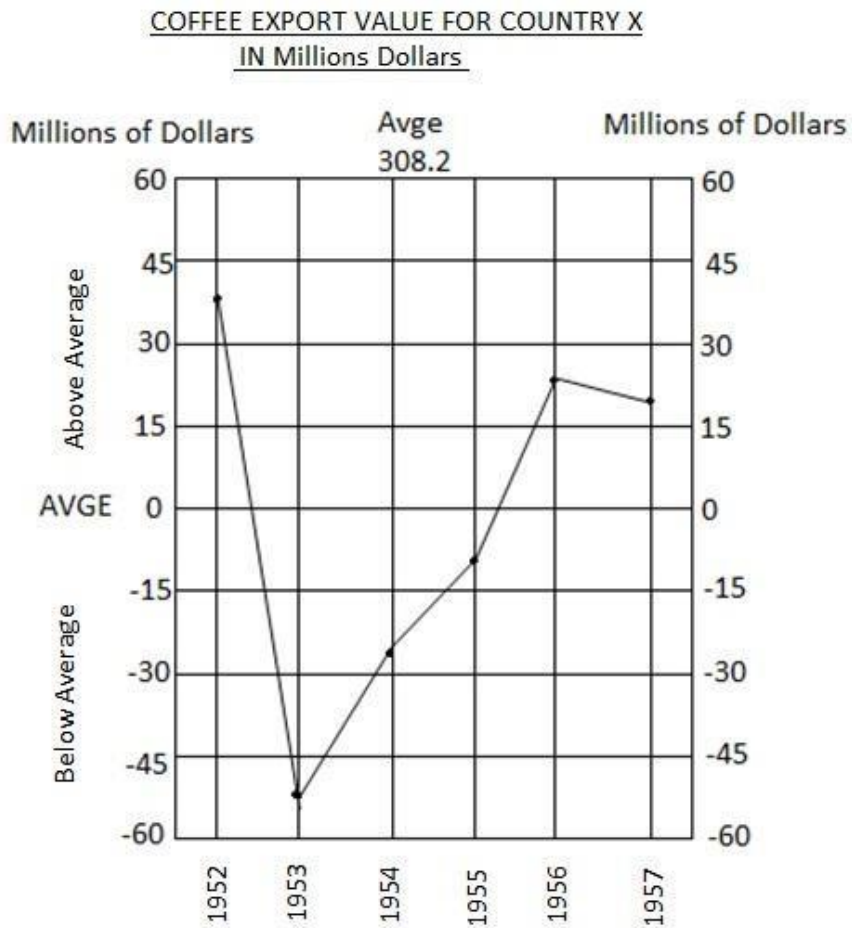
$$\text{Vertical scale} = \frac{\text{Highest deviation value}}{\text{Graph space}}$$

$$\frac{52.5}{4} = 13 - 15$$

Thus: the vertical scale

1cm represents 15 or -15 million dollars

(d) The graph has to be redrawn accordingly as follows:-



Source:-

Hypothetical data

Scales:-

Vertical scale 1cm represents 15 or 15 tons

Horizontal scale 1cm represents 1 year

Merits of the divergent line graph

- The graphical method is useful for showing increase and decrease of the values.
- The graphical method shows the average of all values
- It has perfect replacement by divergent bar graph

Setbacks of the divergent line graph

The graphical method is not suited to show the increase and decrease values for more than one items, it is thus; the graphical method is limited to a single item.

It needs high skill to reveal the actual values of the item represented.

It is time consuming graphical method as its preparation involves a lot of mathematical works.

It requires high skill to construct the divergent line graph.

Group line graph

It is a form of statistical line graph designed to have more than one lines of varied textures to illustrate the values of more than one items. Group line graph is alternatively known as composite, comparative, and multiple line graph.

Construction of the group line graph

Consider the given data below showing values of export crops from Kenya (Ksh Million).

Crop/Year	1997	1998	1999	2000	2001
Tea	24,126	32,971	33,065	35150	34,448
Coffee	16,856	12,817	12,029	11,707	7,460
Horticulture	13,752	14,938	17,641	21,216	19,846
Tobacco	1,725	1,607	1,554	2,167	2,887

(a) Variables identification

Dependent variable export values

Independent variable Date (years)

Y - -axis..... export values

X – axis.....Years

(b) Verticals identification

Dependent variable.....export values

Independent variable Date (Years)

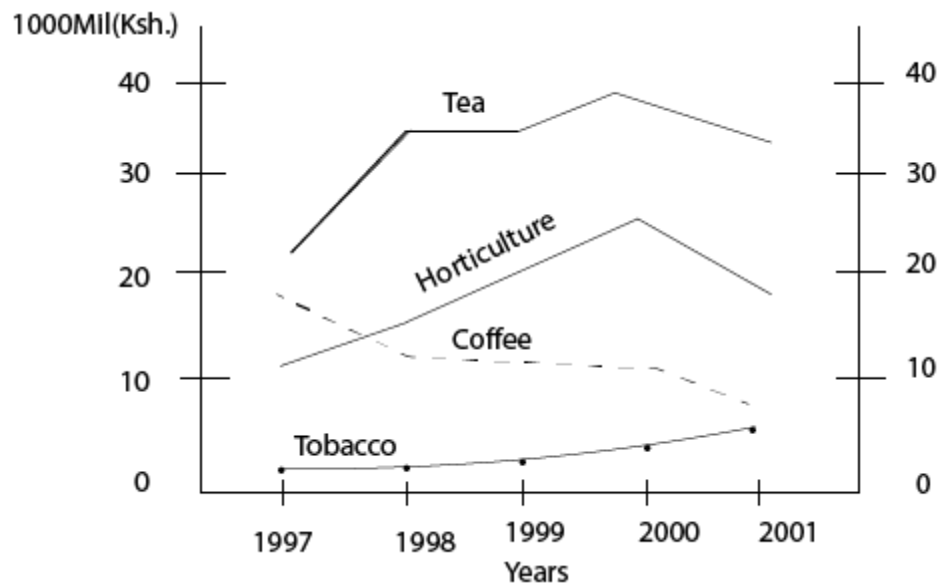
$$\text{Vertical scale} = \frac{\text{Highest value}}{\text{Graph space}}$$

$$\frac{35150}{8 \text{ cm}} = 4393.75 - 5000$$

Hence; VS 1cm represents 5000 export value

Thus; the group line graph appears as follows:-

KENYA: CROPS EXPORT VALUES



Scales:-

Vertical scale: 1cm to 5,000 export values

Source: Kenya Economic Survey 1969

Strengths of the group line graph

- It is much easier to prepare as it involves no complicated mathematical works
- It is useful graphical method for showing the values of more than one cases.

From the graph, the absolute values are extracted as the values directly shown ·
 It is comparatively easier to read and interpret the values. · It has perfect replacement by group bar graph.

Setbacks of the group line graph

- Some times; it becomes difficult to assess the vertical scale if the variation between the highest and lowest values appears wider enough
- Crossing of the lines on the graph may confuse the interpreter.
- A problem may arise in the selection of the varied line textures.

Compound line graph

It is a line graph designed to have more than one lines compounded to one another by varied shade textures to show the cumulative values of more than one items.

Construction of the compound line graph

Consider the given data below showing cocoa production for the Ghana provinces in 000 tons.

YEAR/PROV	TV Togoland	E. province	W. province	Ashanti
1947/48	40	40	30	35
1948/49	50	60	45	100
1949/50	45	46	89	110
1950/51	45	47	44	124
1951/52	47	23	50	100
1952/53	51	14	57	118

Procedure

(a) Variables identification

Dependent variable..... export values

Independent variable Date (Years)

Y - -axis.....export values

X – axis.....Years

(b) Cumulative values determination for the dates.

$$1947/48 \ 40+40+30+35 = 145$$

$$1948/49 \ 50+60+45+100 = 225$$

$$1949/50 \ 45+46+89+110 = 290$$

$$1950/51 \ 45+47+44+124 = 260$$

$$1951/52 \ 47+23+50+100 = 220$$

$$1952/53 \ 51+14+57+118 = 240$$

(c) Vertical and horizontal scales determination

$$\text{Vertical scale} = \frac{\text{Highest commulative value}}{\text{Graph space}}$$

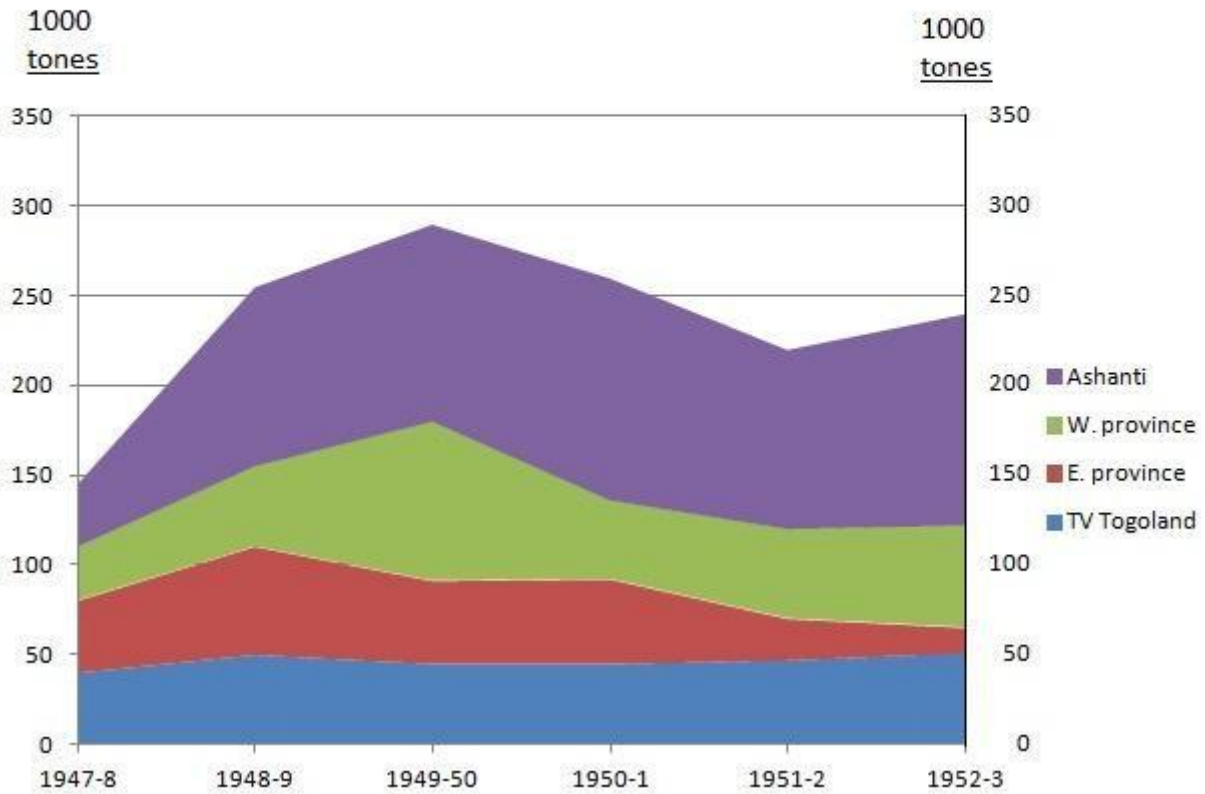
$$\frac{290}{6 \text{ cm}} = 48.3 \approx 50$$

Hence; The vertical scale, 1cm represent 50 tons

Thus the graph appear as follow:-

COCOA PURCHASE FOR GHANA PROVINCES

(1947/48-1952/53)



Strengths of the compound line graph

- It is a useful graphical method for showing the cumulative values of more than one case.
- Depending on the skill the interpreter has, from the graph, the absolute values are extracted as the value directly shown.
- It has perfect replacement by compound bar graph
- It is comparatively easier to assess the vertical scale to be used.

Setbacks of the compound line graph

- It needs high skill to interpret the graph
- It needs high skill to construct the graph
- A problem may arise in the selection of the varied line textures.

BAR GRAPHS

These are the graphs which use bars to illustrate the values of items to give quantitative analysis. Any bar graph has two axes

- X-axis; This is also known as the base or horizontal axis. It is used principally to show the values of independent variable like date or places.

- Y – axis; This is also known as the vertical axis. It is used to show the values for the dependent variable of like output of crops, minerals etc.

TYPES OF BAR GRAPHS

Like line graphs, bar graphs are also extremely varied as differently designed to meet varied functions. With respect to this consideration, bar graphs are categorized into the following:-

- Simple bar graph
- Divergent bar graph
- Group bar graph
- Compound bar graph
- Percentage bar graph
- Population pyramid

Simple bar graph

It is a form of bar graph, designed to have bars of similar texture to illustrate the values of one item in relation to dependent and independent variables. i.e. It is designed to show the values of one item per varied date or places.

Construction of the simple bar graph

Consider the given data below showing cocoa purchase by areas, in 000 metric tons (1953)

Province	Purchase
Ashanti	104
W-Province	39
E-Province	45
TV Togo land	22

Procedures

(a) Variable identification

Dependent variable Purchase

Independent variable Provinces

Y - -axis.....purchase values

X – axis.....Provinces

(b) Verticals identification

Dependent variable.....export values

Independent variable Date (Years)

$$\text{Temperature value scale} = \frac{\text{Highest value}}{\text{Graph pace}}$$

$$\frac{104}{6} = 17.3 \sim 20$$

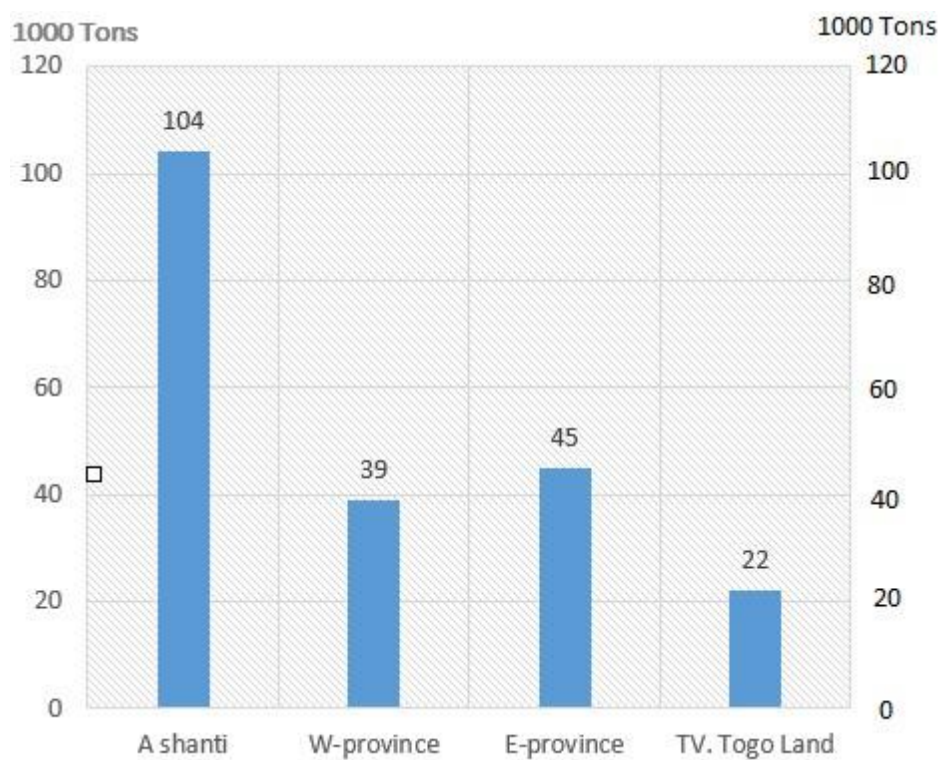
Thus; the vertical scale: 1cm represents 20,000 tons.

Bar width – 1cm

Bar space = 0.5 cm

The graph has to be constructed accordingly.

COCOA PURCHASE BY PROVINCES (1953/54)



Vertical scale; 1cm represents 20000 tons.

Strengths of the simple bar graph

- It is much easier to prepare as it involves no complicated mathematical works, and also bars of similar texture established in the graph.
 - From the graph, the absolute values are extracted. ·
- It is comparatively easier to read and interpret the values ·
- It has perfect replacement by simple line graph.

Setbacks of the simple bar graph

- It is a limited graphical method as only suited to represent the values for one item
- Some times; it becomes difficult to assess the vertical scale if the variation between the highest and lowest values appear wider enough.

Divergent bar graph

It is a form of bar graph designed to illustrate the increase and decrease of the distribution values in relation to the mean. The graph is designed to have **upper and lower sections** showing **positive and negative values** respectively. The two portions are separated by the steady line graduated with zero value along the vertical line. The steady line also shows the average of all values.

Construction of the divergent line graph

Consider the following tabled data which show export values of coffee for country X in millions of dollars.

YEAR	EXPORT VALUES (000,000 dollars)
1952	345
1953	256.5
1954	283
1955	300
1956	335
1957	330.5

(a) Variable identification

Dependent variable Export values

Independent variable Date (Years)

Y - axis..... Export values

X – axis.....Years

(b) Computation of the arithmetic mean

$$\text{Arithmetic mean} = \frac{\sum x}{N}$$

$$\sum x = x_1 + x_2 + x_3 + \text{---} x_n$$

$$345 + 256 + 283 + 300 + 335 + 330.5 + 1850$$

$$\text{Then, } \frac{1850}{6} = 308$$

(c) Computation of the deviation values

$$1952 \ 345 - 308 = 37$$

$$1953 \ 256.5 - 308 = 52.5$$

$$1954 \ 283 - 308 = -25$$

$$1955 \ 300 - 308 = -8$$

$$1956 \ 335 - 308 = 27$$

$$1957 \ 330.5 - 308 = 22.5$$

(d) Estimation of the vertical scale

$$\text{Value Scale} = \frac{\text{Highest Deviation value}}{\text{Graph space}}$$

Thus: the vertical scale 1cm represents 15 or –15 million dollars

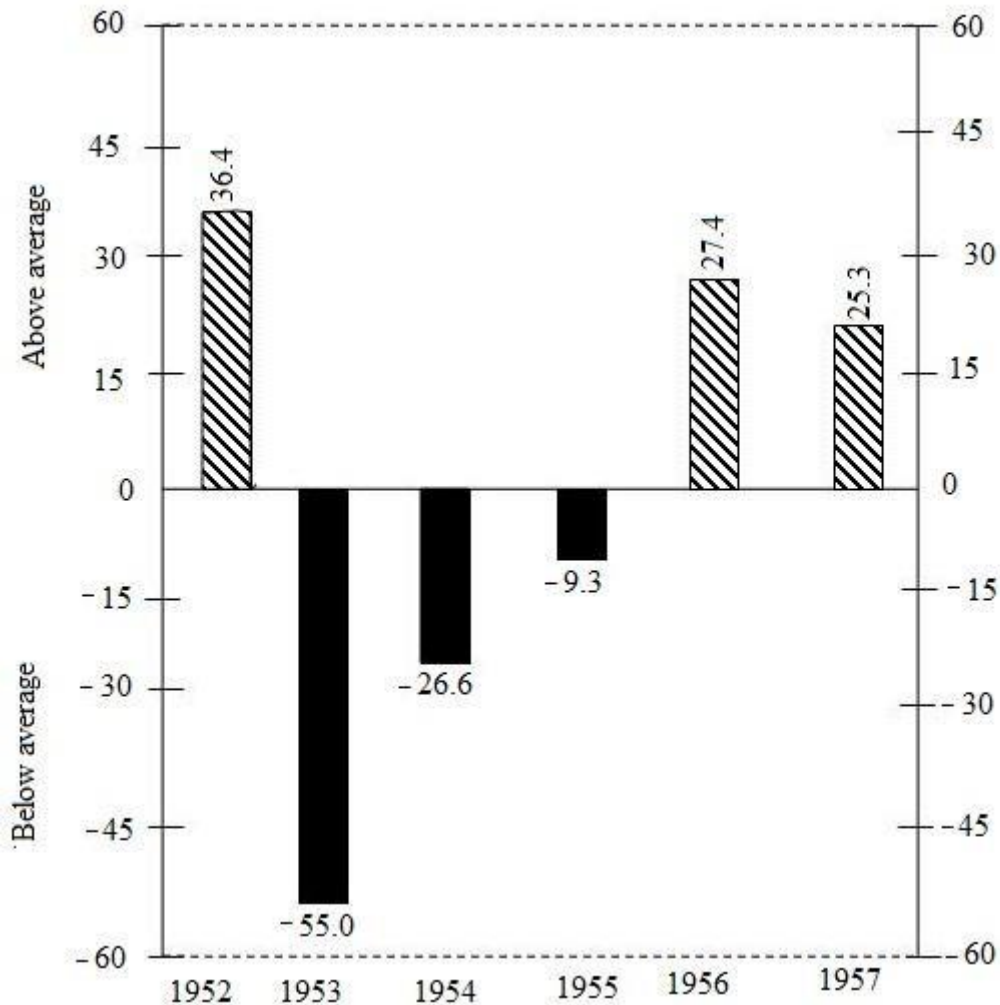
Bar width – 1cm

Bar space – 1cm

(e) The graph has to be redrawn accordingly as follows.

COFFEE EXPORT VALUES FOR COUNTRY X

In million dollars



Scales:-

Vertical scale 1cm represents 15 or – 15 tons

Horizontal scale: 1cm represents 1 year

Source:- Hypothetical data

Merits of the divergent bar graph

- The graphical method is useful for showing increase and decrease of the values
- The graphical method shows the average of all values · It has perfect replacement by divergent line graph.

Setbacks of the divergent bar graph

- The graphical method is not suited to show the increase and decrease values for more than one item, it is thus; the graphical method is limited to a single item.
- It needs high skill to reveal the actual values of the item represented.

- It is time consuming graphical method as its preparation involves a lot of mathematical work.
- It requires high skill to construct the divergent bar graph.

Grouped bar graph

It is a form of statistical bar graph designed to have more than one bars of varied textures to illustrate the values of more than one items.

Grouped bar graph is alternatively known as composite, comparative, and multiple bar graph.

Construction of the group bar graph

Consider the given data below for cocoa purchase by provinces in Ghana (1947/48 – 1950/51)

YEAR/PROV	TV Togoland	E. province	W. province	Ashanti
1947/48	20	54	28	106
1948/49	26	80	46	126
1949/50	24	67	40	116
1950/51	22	72	45	123

(a) Variables identification

Dependent variable..... purchase values

Independent variable Date

Y - -axis..... .purchase values

X – axis.....Date

(b) Vertical scale estimation

$$\text{Vertical scale} = \frac{\text{Highest value}}{\text{Graph space}}$$

$$\frac{126}{7 \text{ cm}} = 18 - 20$$

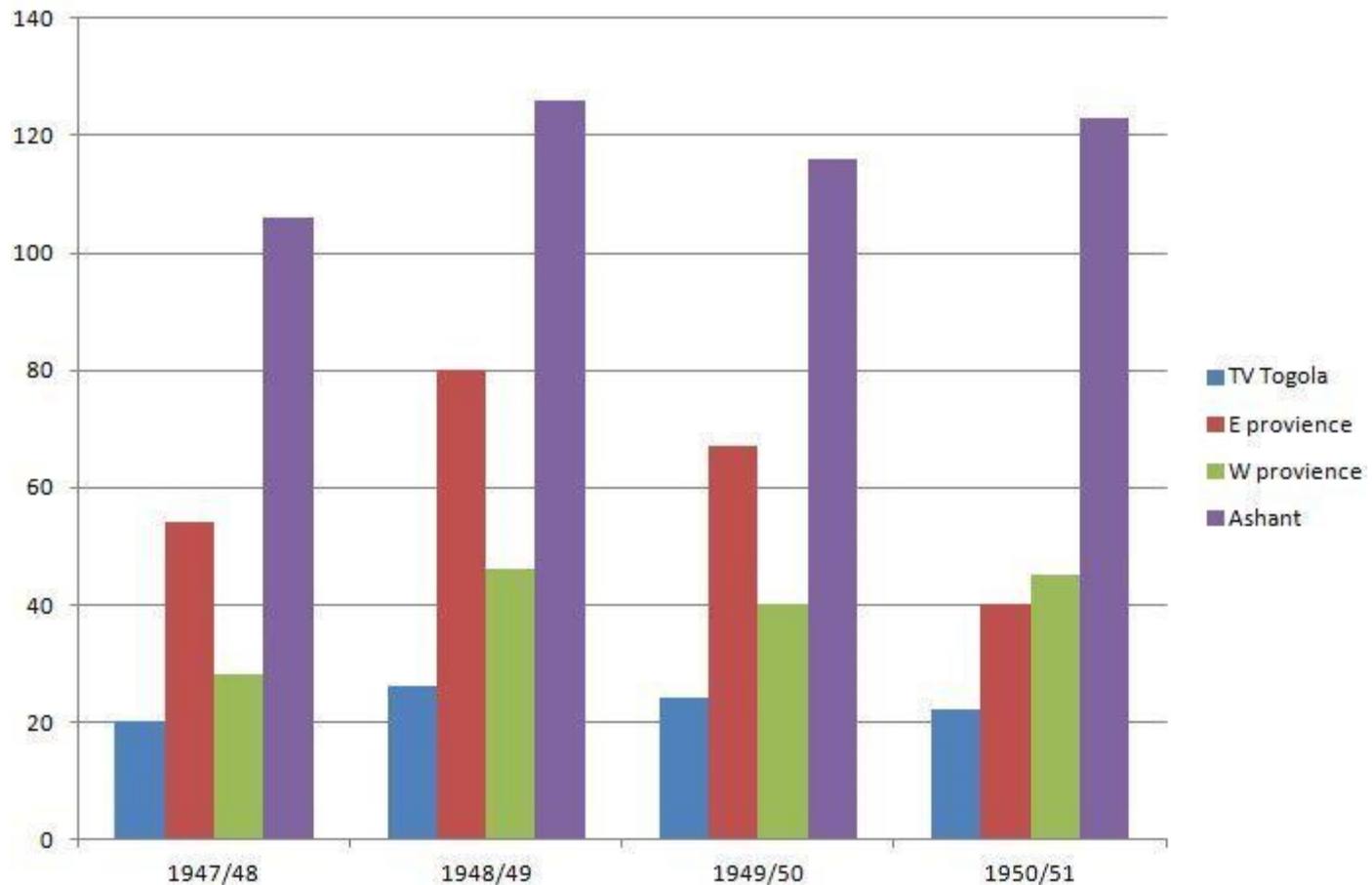
Hence; Vs, 1cm to 20,000 tons

Bar width = 1cm

Bar space = 1cm

(c) The graph should be drawn accordingly.

COCOA PURCHASE BY PROVINCES (1953/54)



Strengths of the grouped bar graph

- It is much easier to prepare as it involves no complicated mathematical works ·
- It is useful graphical method for showing the values of more than one cases. ·
- From the graph, the absolute values are extracted as the value are directly shown ·
- It is comparatively easier to read and interpret the values. · It has perfect replacement by group line graph.

Setbacks of the grouped graph

- Some times; it becomes difficult to assess the vertical scale if the variation between the highest and lowest values appear wider enough.
- A problem may arise in the selection of the varied bar textures.

Compound Bar graph

It is a bar graph designed to have bars divided proportionally showing the cumulative values of more than one items per varied dates or places

Compound bar graph is alternatively known as **divided bar graph**, or **superimposed bar graph**.

Construction of the compound bar graph

Consider the given data below showing cocoa production for the Ghana provinces in 000 tons.

Consider the given data below showing cocoa purchase by provinces (1947/48 to 1950/51)

REGION/YEAR	1947/48	1948/49	1949/50	1950/51
Ashanti	106,000	126,000	116,000	123,000
W.province	28,000	46,000	40,000	45,000
E.Province	54,000	80,000	67,000	72,000
T.Volta	20,000	26,000	24,000	22,000

Procedure

(a) Variable identification

Dependent variable export values

Independent variable ... Date (Years).

Y - -axis..... purchase values

X – axis.....Years

(b) Cumulative values determination for the dates.

$$1947/48..... 106,000 + 28,000 + 54,000 + 20,000 = 208$$

$$1948/49..... 126,000 + 46,000 + 80,000 + 26,000 = 278,000$$

$$1949/50 116,000 + 40,000 + 67,000 + 24,000 = 247,000$$

$$1950/51.....123,000 + 45,000 + 72,000 + 22,000 = 262,000$$

(c) Vertical scale determination.

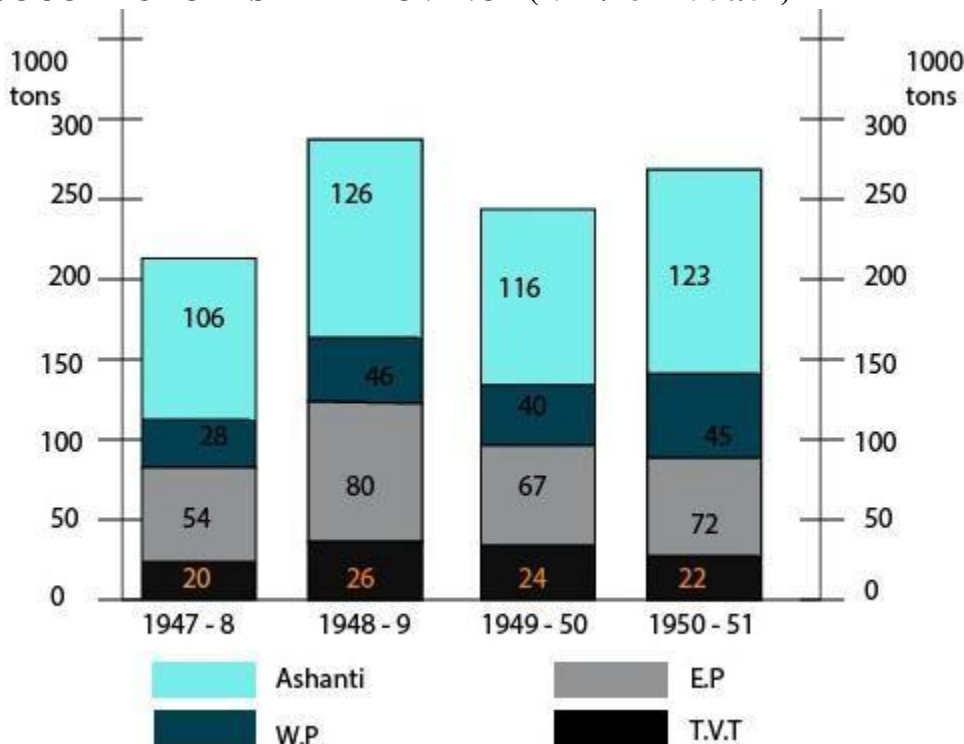
$$\text{Value Scale} = \frac{\text{Highest Cumulative value}}{\text{Graph space}}$$

$$\frac{278,000}{6 \text{ cm}} = 463,000 - 50,000$$

Thus; the VS ... 1cm represents 50,000 tons.

The graph should be drawn accordingly.

COCOA PURCHASE BY PROVINCE (1947/48 – 1950/51)



Strength of the compound bar graph

- It is useful graphical method for showing the cumulative values of more than one cases
- Depending on the skill the interpreter has, from the graph, the absolute values are extracted as the value directly shown.
- It has perfect replacement by compound line graph
- It is comparatively easier to assess the vertical scale to be used.

Setbacks of the compound bar graph

- It needs high skill to interpret the graph
- It needs high skill to construct the graph
- A problem may arise in the selection of the varied textures of the proportional segments

It is very tedious /tiresome as it involves mathematical calculation
It is time consuming in preparation

Percentage bar graph

In percentage bar graph, all bars must be drawn on the same height representing 100% and a suitable scale is chosen such as 5, 10, 20 etc, and marked along the sides. The percentages of the total each area stands for must start from zero line. Also it is advised to include the actual percentages of the face of the bars.

Construction of the Percentage bar graph

Consider the given data below showing cocoa purchase by provinces (1947/48 to 1950/51)

REGION/YEAR	1947/48	1948/49	1949/50	1950/51
Ashanti	106,000	126,000	116,000	123,000
W.province	28,000	46,000	40,000	45,000
E.Province	54,000	80,000	67,000	72,000
T.Volta	20,000	26,000	24,000	22,000

Procedure

(a) Variables identification

Dependent variable export values

Independent variable ... Date (Years).

Y-axis..... purchase values

X-axis.....Years

(b) Cumulative values determination for the dates.

1947/48..... $106,000 + 28,000 + 54,000 + 20,000 = 208$

1948/49..... $126,000 + 46,000 + 80,000 + 26,000 = 278,000$

1949/50 $116,000 + 40,000 + 67,000 + 24,000 = 247,000$

1950/51..... $123,000 + 45,000 + 72,000 + 22,000 = 262,000$

(c) The percentages by provinces in each year determination. **1947/48:**

$$\text{Ashanti : } \frac{106,000}{208,000} \times 100\% = 51\%$$

$$\text{W. province : } \frac{28,000}{208,000} \times 100\% = 13.4\%$$

$$\text{E. province: } \frac{54,000}{208,000} \times 100\% = 26\%$$

$$\text{Tv/T : } \frac{20,000}{208,000} \times 100\% = 9.6\%$$

1948/49:

$$\text{Ashanti:} = \frac{126,000}{278,000} \times 100\% = 45.3\%$$

$$\text{W. Province:} = \frac{46,000}{278,000} \times 100\% = 16.5\%$$

$$\text{E. Province:} = \frac{80,000}{278,000} \times 100\% = 28.77\%$$

$$\text{Tv/T:} = \frac{26,000}{278,000} \times 100\% = 9.4\%$$

1949/50:

$$\text{Ashanti:} = \frac{116,000}{247,000} \times 100\% = 47\%$$

$$\text{W. Province:} = \frac{40,000}{247,000} \times 100\% = 16.2\%$$

$$\text{E. Province:} = \frac{67,000}{247,000} \times 100\% = 27.1\%$$

$$\text{Tv/T:} = \frac{24,000}{247,000} \times 100\% = 9.7\%$$

1950/51:

$$\text{Ashanti:} = \frac{123,000}{262,000} \times 100\% = 46.94\%$$

$$\text{W. Province:} = \frac{45,000}{262,000} \times 100\% = 17.17\%$$

$$\text{E. Province:} = \frac{72,000}{262,000} \times 100\% = 27.48\%$$

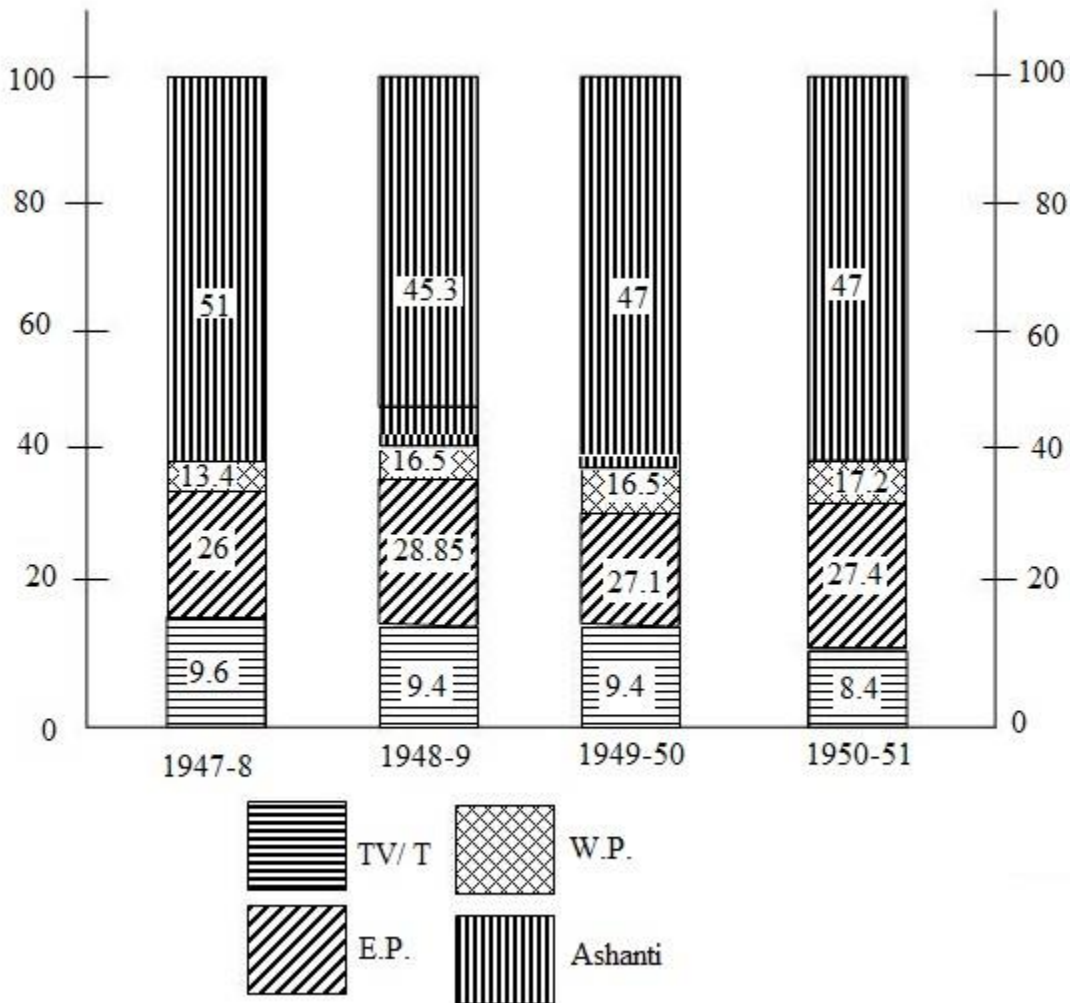
$$\text{Tv/T:} = \frac{22,000}{262,000} \times 100\% = 8.4\%$$

$$\text{Vertical scale} := \frac{100}{\text{Graph space}}$$

$$\frac{100\%}{5 \text{ cm}} = 20\%$$

Hence; VS; 1 cm represents 20%

The percentage bar graph should be drawn accordingly as follow:- **COCOA PURCHASE BY PROVINCES (1947/48 – 1950/51)**



Vertical scale; 1cm represents 20%

Strengths of the percentage bar graph

- It is useful graphical method for showing the values of more than one cases
- The data represented appear in a more simplified form as given in percentages.
- It is comparatively easier to assess the vertical scale to be used

Setbacks of the percentage bar graph

- It does not give the absolute values

- It needs high skill to interpret the graph
- It needs high skill to construct the graph
- A problem may arise in the selection of the varied textures of the proportional segments
- It consumes much time to be prepared.

Population pyramid graph

- It is a form of bar graph designed to show population distribution by age and sex. **It is a double bar chart showing the age sex structure of the population.** It consists of two sets of horizontal bars; one is for each sex showing either the percentages or absolute numbers.

Rules for drawing the population pyramid graph

- It is a principle in drawing population pyramid; the number of male population illustrated by the left set of bars; while that of females by the right set of bars.
- The young population distribution is always at the bottom while that of old at the top.
- Usually the last age group should be left open handled because; some people may survive beyond 100 years and their number have been omitted.
- The bottom scale can be graduated as percentages or absolute numbers.
- If percentages are opted to be used; the total population of both combined sexes should be used to compute the percentages.
- After all the bars have been drawn, they can be shaded in one colour or separated colours for each sex.

CONSTRUCTION OF THE POPULATION PYRAMID

There are two techniques of drawing the horizontal bars of an age sex pyramid. In the first technique, the bars are drawn proportionally to the actual population numbers (absolute values).

In the second technique, the bars are drawn to represent percentages.

Age group	Male	Female	Total
0 – 4	2291936	2242966	4534902
5-9	2000580	1962556	3963136
10-14	2034980	2003655	4038635
15-19	1681984	1721194	3403178
20-24	1328529	1504389	2832918
25-29	1094909	11664594	2259503
30-34	840692	845230	1685922

35-39	695263	723749	1419012
40-44	516502	516989	1033491
45-49	419841	418987	838828
50-54	344639	340167	684806
55-59	223691	236325	460016
60-64	194513	214715	409228
65-69	140969	160364	301333
70-74	118601	135524	254125
75-79	79166	81620	160786
80+	95300	121038	216338
Age not stated	103487	86956	190443
All ages	14205589	14481018	28686607

The absolute value technique

The following steps are followed when constructing a population pyramid using absolute values.

- Decide a suitable scale for the horizontal axis (baseline) by considering the values of the biggest and smallest age group, as well as the size of the paper on which the pyramid is to be drawn. Horizontal scale is determined as follows.

$$\text{Horizontal scale} = \frac{\text{Highest value}}{\text{Graph space}}$$

$$\frac{2291936}{6} = 381989.3 - 400,000$$

Hence by considering the data in the table, scale of 1cm to represent 400,000 people would be suitable.

- Choose a suitable scale for the vertical axis. This scale will determine how wide the bars will be and also the interval between the age groups. The width of the bars should not exceed 6mm otherwise the pyramid will look untidy.
- Take a clean graph paper and on it draw horizontal axis at least 3 cm from the bottom of the page. Draw two vertical axes of 1 cm apart and about 10 cm long, until they touch the horizontal axis.
- Where the vertical axes touch the horizontal axis, mark as zero. On the horizontal axis, and at intervals of 1cm from the zero mark on the both sides, mark of the values representing the female and male populations.

- In the middle column, fill in the age groups starting with the youngest at the bottom. The age groups should be within the width of the horizontal bars.
- Using the horizontal scale, and starting with the first age group for females, draw a bar from the vertical axis on the right hand side of the central column towards the right to represent the female population of that group. The scale chosen in step 1 above will determine the length of the bar.
- From the left hand side of the vertical axis, draw a bar representing the male population of the same age group. Steps 6 and 7 should be repeated for all the subsequent age group until the last one has been represented.

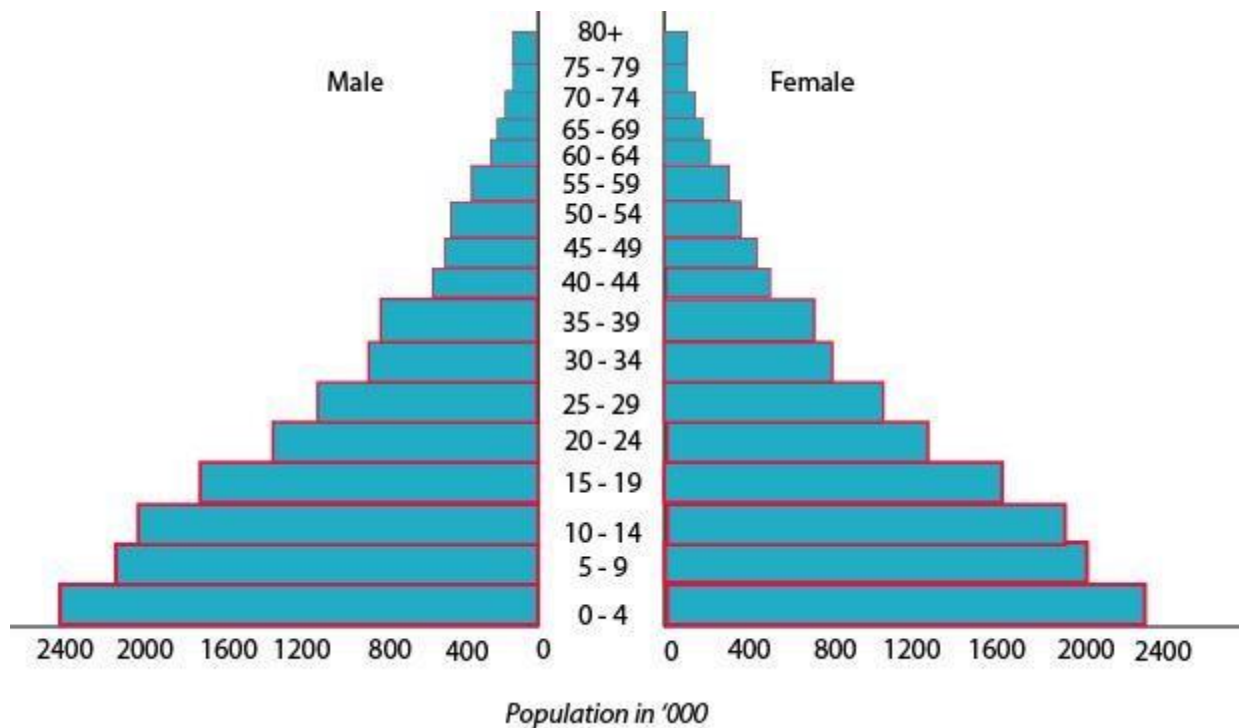


Fig. 1.1 Kenya: Population by age and sex, 1999

The percentages technique

By this technique, the values for population distribution by age and sex given in percentages.

The percentages of each female or male group over the total populations is calculated from the absolute values in our example and a new set of data will be derived from data in the table. This new data will be used to draw the graph.

An example on how to calculate the percentage values is shown below. The application for calculating the percentage is as follows.

$$\text{percentage value} = \frac{\text{Population}}{\text{Total population size}} \times 100\%$$

For instance:

The absolute values for the females aged between 0-4 years from the table is 2242 966, while that for males is 2291936. The total populations according to the 1999 census, was 28686607. Therefore the percentage of females is as follows:-

$$\frac{2242966}{28686607} \times 100 = 7.8\%$$

The percentage of male is as follows:-

$$\frac{2291936}{28686607} \times 100 = 8.0\%$$

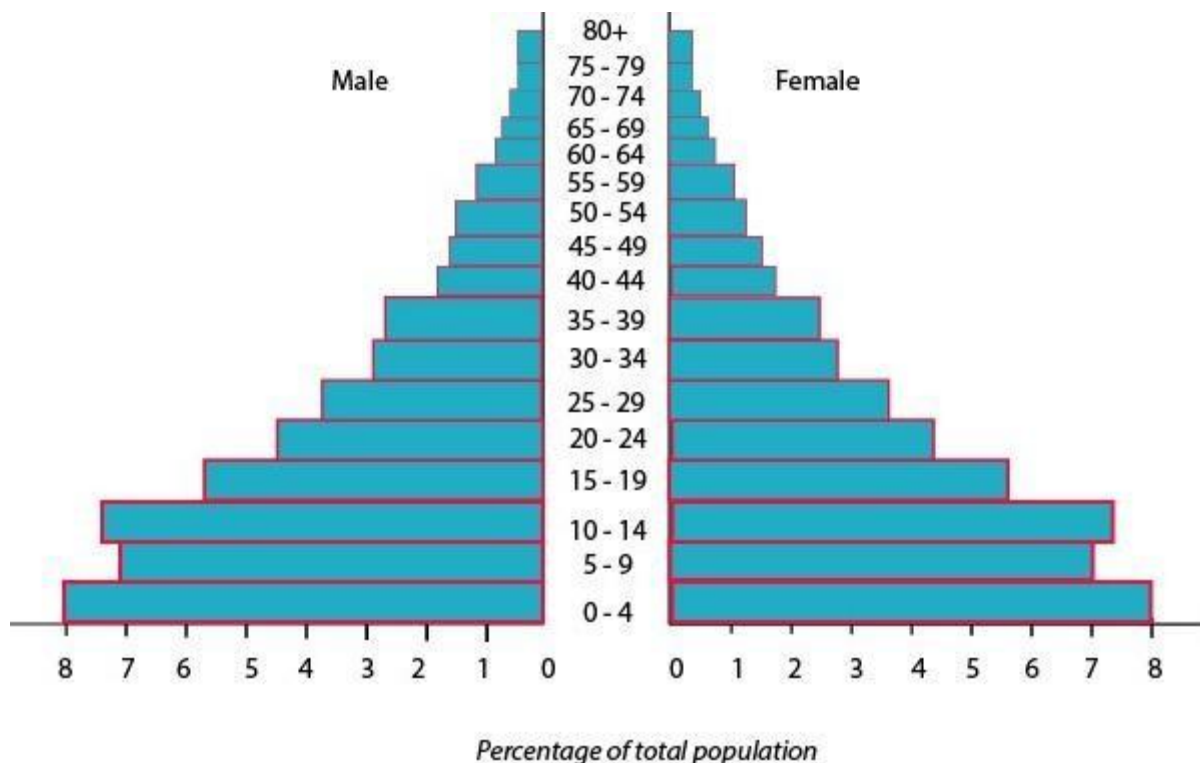
The worked out percentage values from the figure in the table are given in the table next page.

Age Group	5male	%female	Total
1-4	8.0	7.8	15.8
5-9	7.0	6.8	13.8
10-14	7.1	7.0	14.1
15-19	5.9	6.0	11.9
20-24	4.6	5.2	9.8
25-29	3.8	4.1	7.9
30-34	2.9	2.9	5.8
35-39	2.4	2.5	4.9
40-44	1.8	1.8	3.6
45-49	1.5	1.5	3.0
50-54	1.2	1.2	2.4
55-59	0.8	0.8	1.6
60-64	0.7	0.7	1.4
65-69	0.5	0.6	1.1
70-74	0.4	0.5	0.9
75-79	0.3	0.3	0.6
80+	0.3	0.4	0.7

After the calculation of the percentages, the following steps should be taken to come up with the age – sex pyramid.

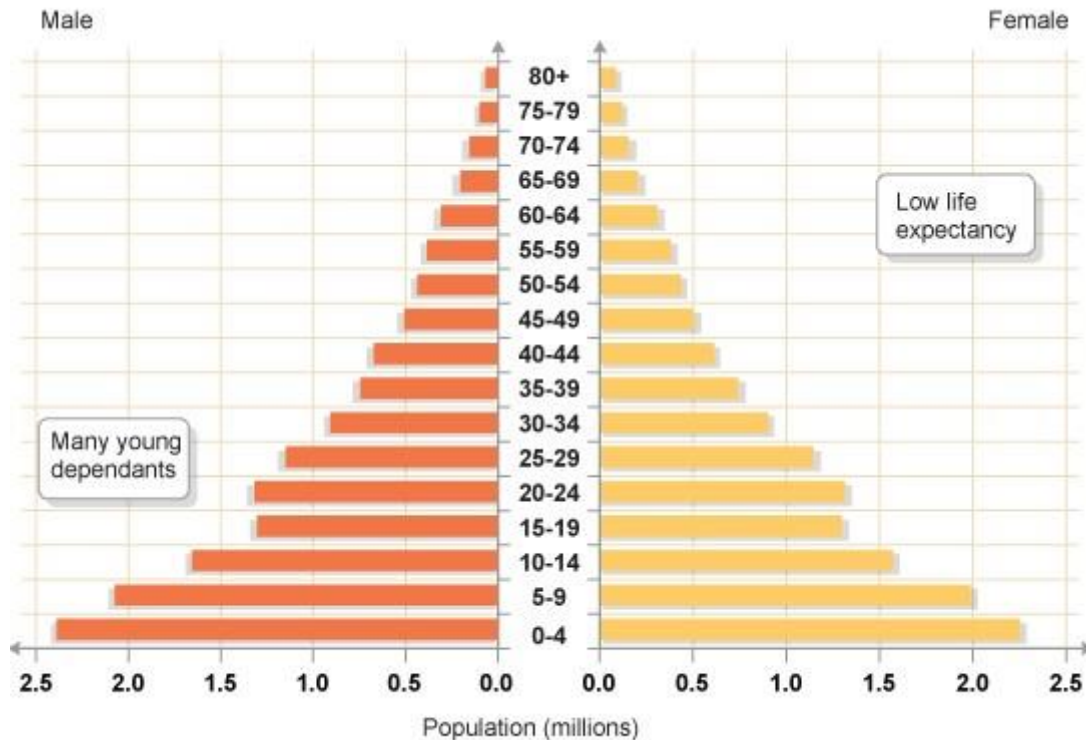
- Choose a suitable scale for the horizontal axis by considering the highest and the lowest percentages in the table. According to the values in the table, a scale of 1cm is representing 1% would be suitable.
- Follow step 2 and 3 as outlined under the absolute values techniques discussed earlier.
- Where the vertical axis touches the horizontal axis, mark zero and at intervals of 1cm, mark off the percentage value towards the right for females, and towards left for the males.
- The age group should be indicated in the middle column just as we did when constructing an age sex pyramid using absolute values..
- Using the horizontal scale and starting with age group 0-4 draw a bar on the right hand side to represent the percentage values of the female population in this age group. In our example, the percentage is 7.8 Draw a similar bar on the left hand side to represent the value of the male population, which in our case is 0.8.
- Draw bars to represent all the age groups follow steps 9 and 10 under the absolute value technique to complete the pyramid.

Kenya population by age:



Note

Pyramid may also be for the purpose of making comparison either in terms of time or location. This can be by means of a double combined population pyramid. The double combined population pyramid looks as follows.



Advantages of the age-sex pyramid

- It is visually attractive method of presenting data.
- A variety of information is shown on the same graph. The details include; age, sex and number of people.
- It can be used to compare the age sex structure of number of countries
- It gives a clear picture and summary of the population composition of a country.

Disadvantages of the age-sex pyramid

- It is tedious to construct because it involves many values.
- It is difficult to tell the exact values represented because of the small scale of the horizontal axis.
- Reasons for the differences in population numbers cannot be obtained from the graph directly. Therefore additional information has to be thought from elsewhere.

COMBINED BAR AND LINE GRAPH

It is a form statistical graph designed to have both bars and line to show two attributes whose values appear in varied unit. It is basically employed to show the values of rainfall and temperature together in a year.

In the graph, the bars used to illustrate the values on amount of rainfall in mm or inch, while the line is used to illustrate the values on amount of temperature in $^{\circ}\text{C}$ or $^{\circ}\text{F}$. This is also known as **climo graph**.

Construction of the bar and line graph

Consider the following climatic data for Dar-win weather station Australia.

Month	J	F	M	A	M	J	J	A	S	O	N	D
Temp °C	28.9	27.8	28.9	29	26.7	26	25.1	26.4	28.1	29.7	29.8	29
Rain(mm)	388	330	246	114	17.8	5	2.5	2.5	12.7	53.3	132	261

Procedure

(i) Identification of the variables

- Dependent variable – Rain and temperature values
- Independent variable – Data (months).

Y - -axis – Rain and temperature values

X – axis.....months

(ii) Estimation of the vertical scale to be used

$$\text{Rainfall vertical scale} = \frac{\text{Highest value}}{\text{Graph space}}$$

$$\frac{388 \text{ cm}}{8 \text{ cm}} = 48.5 - 50 \text{ mm}$$

Thus; the vertical scale for rainfall is 1cm to 50mm.

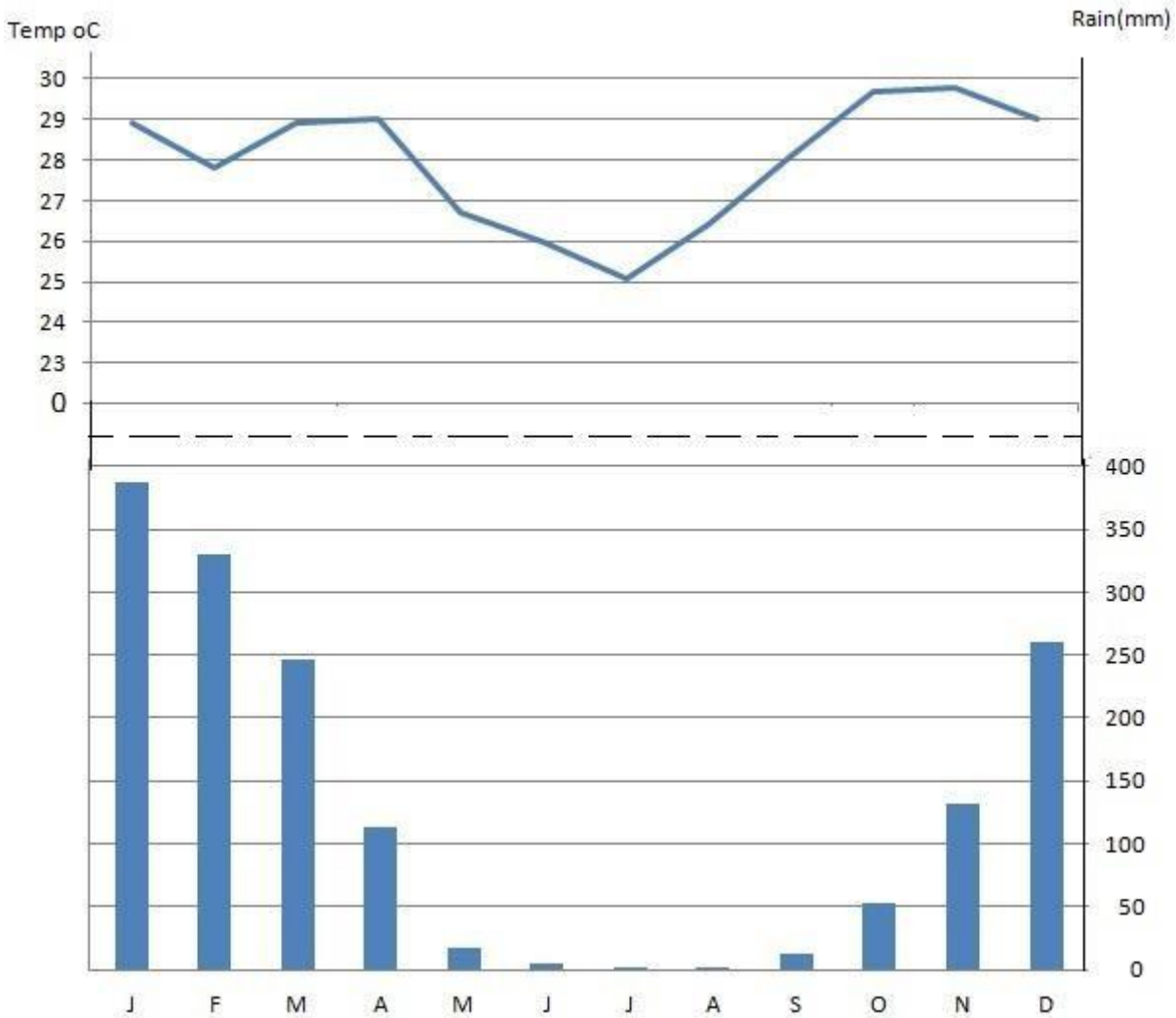
$$\text{Temperature vertical scale} = \frac{\text{Highest value}}{\text{Graph space}}$$

$$\frac{29.8\text{c}}{3 \text{ cm}} = 9.6 - 10$$

Thus; the vertical scale for temperature is 1 cm to 10 c

(iii) The graph has to be drawn as follows;

CLIMATIC CONDITION FOR DARWIN AUSTRALIA



Strengths of the combined bar and line graph

- It is useful graphical method for showing the distribution values of climate
- It is more illustrative, as it provides visual idea to the users in statistics.
- It allows the easy making of quantitative analysis

Setbacks of the combined bar and line graph

- It is more illustrative, as it provides visual idea to the users in statistics
- Needs high skill to make quantitative analysis from the graph
- It is time consuming graphical method in construction
- It needs high skill to construct the graph
- It is tedious as it involves mathematical calculation

STATISTICAL MAPS

As it has been introduced in the chapter one; the numerical data after to have been collected, summarized and analyzed; are presented to provide **pictorial view (visual idea)**. One of the useful ways for representing the numerical facts is by maps. The method of maps is established with an emphasize of showing distribution values of phenomena of places over the earth's surface. Moreover; the places whose values to be shown on maps should lie adjacently to one another in such a way they can all appear on a similar map. It is thus; **statistical maps are the ones designed to show the values on spatial distribution of geographical events (phenomena)**.

Or

The maps designed to show spatial distribution of certain geographical events in quantitative manner and in turn allow quantitative analysis.

The main useful statistical maps which allow quantitative analysis include the following.

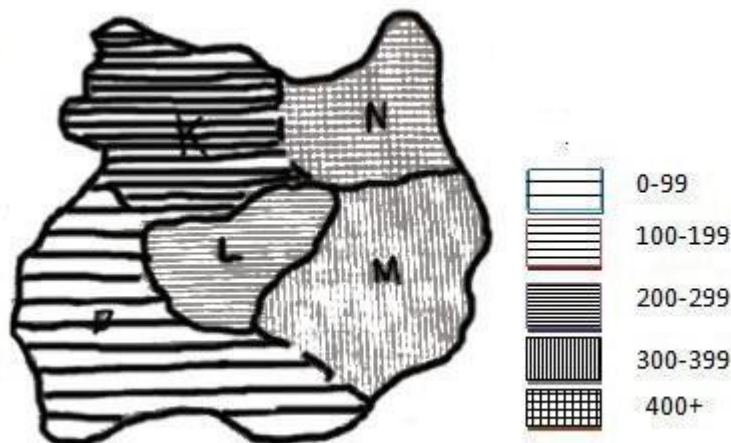
- Choropleth maps
- Dot maps
- Flow line maps ·

Isopleths maps

Choropleth maps

These **are the statistical maps which use the system of varied shade textures to illustrate the density spatial distribution values of a certain phenomena for the places**. The maps mostly designed to show population density of places. On the map, places with similar shade texture have almost the same distribution density.

Eg.



Construction of the choropleth map

- (i) Obtain the base map with suitable scale. The map should have the boundaries of administrative areas. The scale is used to assess the size of the administrative areas which then are related to the amount of distribution.
- (ii) Obtain the data and summarize into the table. The tabled data should show clearly the names of administrative areas, area size and amount of distribution.
- (iii) Determine the density values of distribution for areas
- (iv) The worked densities should be grouped using regular interval. In this respect; more than one classes should be selected and all should have the worked densities. It is also important that; the classes should not be numerous.

Example:-

Use the following data to prepare the choropleth map:-

Province	Population	Land area (km ²)
Nairobi	2,143,254	696
Central	3,724,159	13,220
Coast	2,487,264	82,816
Eastern	4,631,779	153,473
North Eastern	962,143	128,124
Nyanza	4,392,196	12,547
Rift valley	6,987,036	182,539
Western	3,358,776	8,264

Calculation of the population densities for the regions.

Population Density = $\frac{\text{Population size}}{\text{Area size}}$ then :

$$\text{Nairobi} = \frac{2143254}{696} = 3079 \frac{\text{person}}{\text{Km}^2}$$

$$\text{Central} = \frac{3,724,159}{13220} = 282 \frac{\text{person}}{\text{Km}^2}$$

$$\text{Coast} = \frac{2,487,264}{82,816} = 30 \frac{\text{person}}{\text{Km}^2}$$

$$\text{Eastern} = \frac{4,631,779}{153,473} = 30 \frac{\text{person}}{\text{Km}^2}$$

$$\text{Northern Eastern} = \frac{962,143}{128,124} = 8 \frac{\text{person}}{\text{Km}^2}$$

$$\text{Nyanza} = \frac{4392196}{12547} = 350 \frac{\text{person}}{\text{Km}^2}$$

$$\text{Rift valley} = \frac{6,987,036}{182,539} = 38 \frac{\text{person}}{\text{Km}^2}$$

$$\text{Western} = \frac{3,358,776}{8,264} = 406 \frac{\text{person}}{\text{Km}^2}$$

The suggested interval is of 100 and thus; the groups include:- 0-99, 100-199, 200-299, 300-399, 400-499 and 500.

With regards to the groups, the choropleth map appears as follows:

KENYA: DENSITY POPULATION BY PROVINCE 1999



Advantages of the choropleth map

- It is most suited to show distribution values of a certain geographical phenomena in relation to area size i.e. It is the most suited to show densities over space.
- The data can be analyzed quantitatively from the map
- It provides visual idea (impression) to people on varied densities of distribution.

The disadvantages of choropleth map

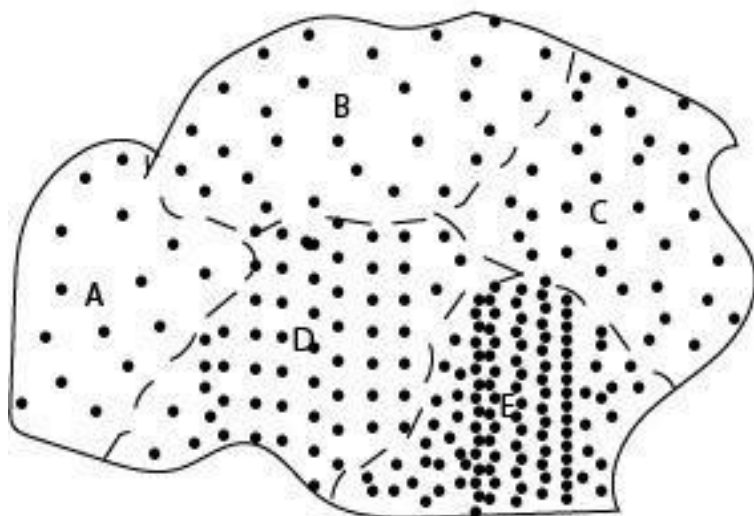
- The shades indicated on the map remove the political boundaries
- It is tedious enough in construction as it involves many values. i.e. preparation consumes much time.
- If no topographical map provided the map may give wrong picture about the distribution of the phenomena.
- Problem may occur in deciding the varied shade textures to be used on the map.
- The map might be realized to show abrupt change of distribution area to area some thing which is not realistic
- It is not possible to obtain absolute values or exact densities from the map because the shades represent categories of densities.
- It is not possible to insert additional details on the map.

Dot maps

Dot map is a **considerable form of statistical map which involves the use of fixed size dots to show the spatial distribution of a certain geographical phenomena like people, cattle etc.**

Or

A map which shows the spatial distribution numerical quantities using dots. A dot is a simplest symbol used in representing quantities on maps. A dot represents fixed amount similarly to others.



Construction of a dot map

- Take into consideration the base map given. The base map should have the clear boundaries of the administrative areas.
- Obtain the data and summarize into a table. The tabled data should show the names of administrative areas and their amount of distribution for the phenomena.
- Make decision on dot value. In this, it is important for the dot value should not be too small or too large. With too large dot value, there is a possibility for the regions with small amount of distribution to lack dots and thus; may impress that, the areas less occupied. If too small dot value chosen, may cause a problem of dots overlapping. It is thus; the dot value should be reasonable.
- Determine the number of dots to be allocated in the administrative areas on the map. It is by dividing the amount of distribution to dot value.
- Insert the dots on the map accordingly. It is important for all dots to have the same size and evenly distributed.

Example:-

Province	Population
Nairobi	2,143,254
Nairobi	2,143,254
Central	3,724,159
Coast	2,487,264
Eastern	4,631,779
North Eastern	962,143

Nyanza	4,392,196
Rift valley	6,987,036
Western	3,358,776

Procedure:-

(i) Dot value determination

According to the given data; 1 dot represents 100,000 people (ii)

Number of dots determination

$$\text{Number of Dots} = \frac{\text{Amount of Distribution}}{\text{Dot value}}$$

$$\text{Nairobi} = \frac{2143254}{100000} = 21 \text{ dots}$$

$$\text{Central} = \frac{3,724,159}{100000} = 37 \text{ dots}$$

$$\text{Coast} = \frac{2,487,264}{100000} = 25 \text{ dots}$$

$$\text{Eastern} = \frac{4,631,779}{100000} = 46 \text{ dots}$$

$$\text{Northern Eastern} = \frac{962,143}{100000} = 10 \text{ dots}$$

$$\text{Nyanza} = \frac{4392196}{100000} = 44 \text{ dots}$$

$$\text{Rift valley} = \frac{6,987,036}{100000} = 70 \text{ dots}$$

$$\text{Western} = \frac{3,358,776}{100000} = 34 \text{ dots}$$

KENYA:POPULATION BY PROVINCES 1999 (* 100,000 people).

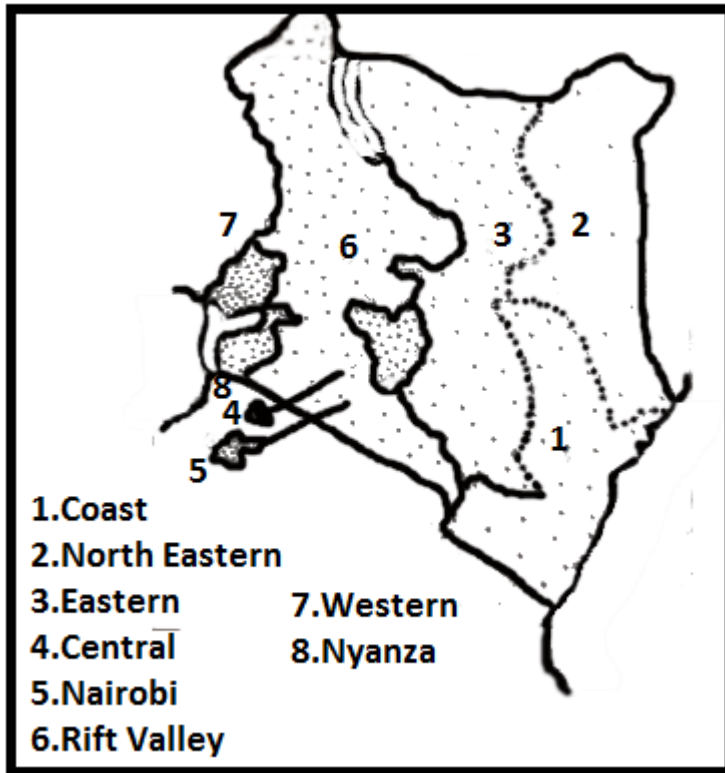


Fig. 1.3 Kenya: Population by provinces, 1999 (100,000 people) **Advantages of dot maps**

- The data can be analyzed quantitatively from the map.
- It is easy to get the amount of distribution of each area by considering the number of dots present and the dot value.
- Preparation of the map is fairly easy
- The map provides visual impression
- They are the most widely used statistical maps for showing distribution.

Disadvantages of dot maps

- The map is facing a problem of double counting during of making quantitative analysis. This give wrong quantitative picture.
- If no topographical map provided, the map may give wrong picture about the distribution of the phenomena
- With larger or smaller dot values, problem may occur in representing distribution on the map.
- Fractional values may not be represented on the map
- Drawing many dots of uniform size is difficult. Special pens may be needed for this purpose.

FLOW LINE MAPS

These are the maps which illustrate the volume of goods or number of vehicles, people, cattle e.t.c. moving between points or areas along established routes of like roads, railways, canals, or air and sea routes.

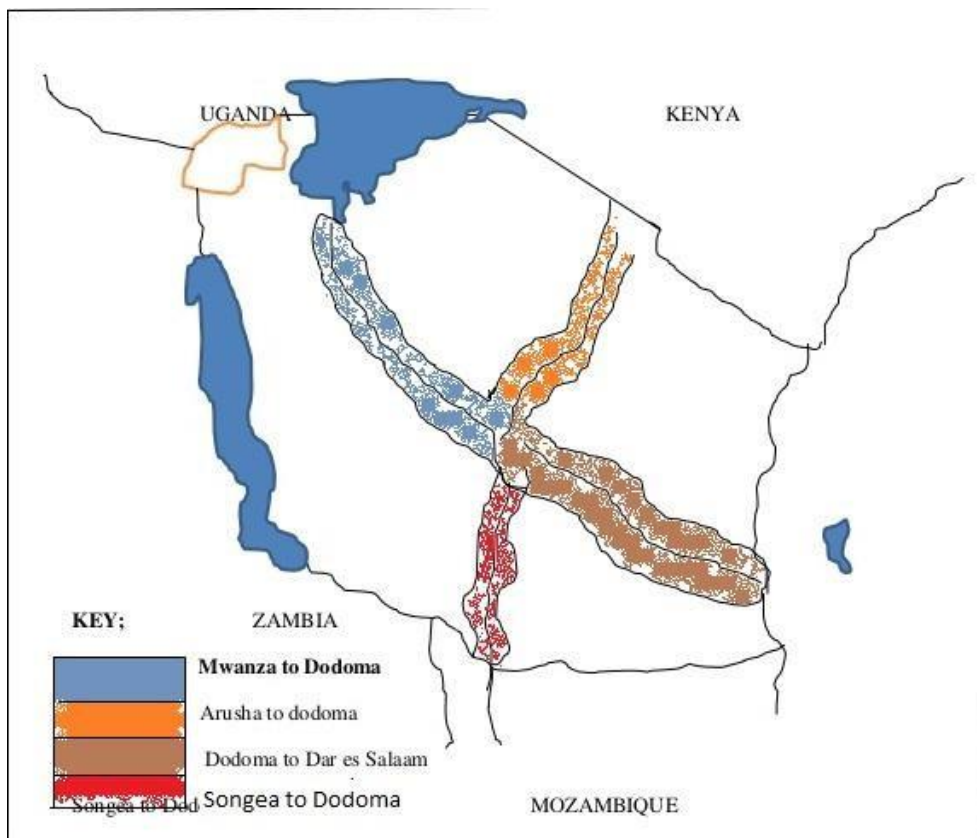
Or

A statistical map designed to show the movement of the geographical phenomena from one place to another through an established route way of like road, railway, water way, airway and others.

With the flow line map, a line shows the direction of the movement; while, the amount of movement is by varied width line. The character of the movement can be by varied shade textures or colours.

It has to be noted that; the direction of the movement and the distance involved have no significance as far as quantities are concern.

E.g.



Construction of the flow line map

- Draw the base map of route ways
- Asses the data given. The data should have names and amount of movement between the check points (stations) along the route way.

- (c) Decide the width scale value. This has to take into consideration the highest and lowest values. It is much better to avoid too large or too small scale values. Too large scale values makes very fine flow lines and too small scale value may result into wider flow lines.
- (d) With respect to the decided scale, draw the flow lines along the routes on the map.

Example:-

Use the data and map given, to show the amount of movement of the passengers between the check points along the route ways.

CHECK POINTS	PASSENGERS
A – B	10,000
B – C	8,000
B – D	7,000
B – E	7,000
E – F	3,000
E – G	2,000

Procedure

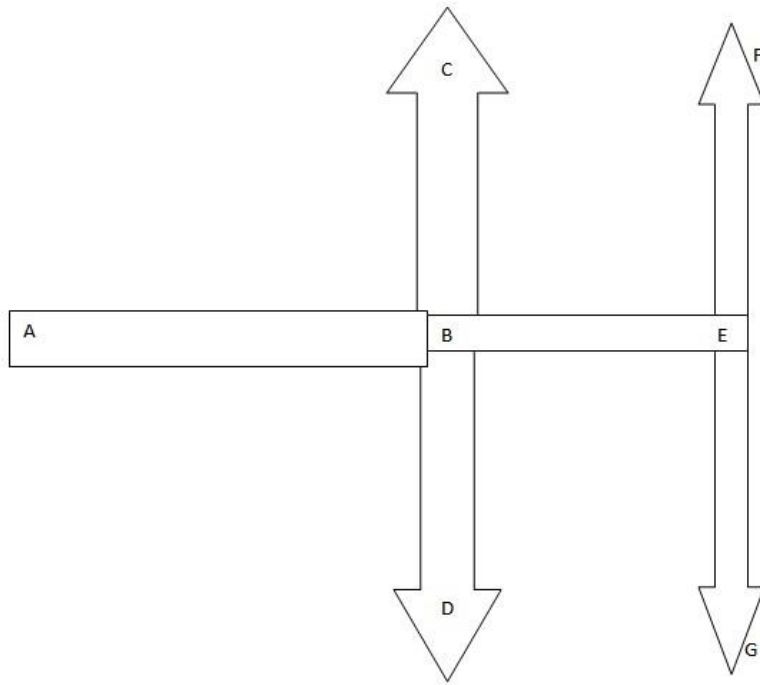
Scale value determination:

$$\text{Value Scale} = \frac{\text{Highest value}}{\text{Highest flow line width}}$$

$$\frac{100000}{1 \text{ cm}} = 100,000$$

Thus; along the flow line; 1mm represents 1,000

The flow lien map for the data given appears as follows.



Advantages of the flow line map

- The map is most useful for showing the amount (volume) of movement between the check points along the route ways.
- The data from the map can be quantitatively analyzed by regarding the width of the flow lines and the value scale.
- It provides visual impression to people
- Calculation and drawing of it is fairly easy once the scale value has been decided.

Disadvantages of the flow line map

- Wide variation between the highest and the lowest values given difficult to assess the scale value
- The volume (amount) of movement cannot be exactly analyzed from the map.
- Difficult may arise in drawing the double flow lines
- The very small values always are not accurately represented on the map.

ISOPLETH MAPS

It is form of statistical map which uses the system of lines to show amount of distribution of phenomena. The lines on the map are drawn to connect points of equal values and the lines are called isolines.

Isopleths maps are also called isoline map, isarithm map and isometric map.

Examples of isopleths maps include; relief map by contours, meteorological maps showing atmospheric pressure, rainfall, temperature, etc. and maps which show depth of water bodies.

The isolines established on the isopleths map have special terms for specialized purposes.

- Isotherms - Temperature
- Isobars – Atmospheric pressure
- Isohyets – Rainfall
- Isoneph – cloudiness
- Isobaths ocean depth
- Isohaline – salinity

Construction of the Isopleths maps

1. Obtain the outline base map and the appropriate data and mark in the points and their values in pencil on the map.
2. Decide the interval to be used
3. Select the critical values. These are the ones which correspond (match) with the chosen interval.
4. Join the critical values with smooth lines according the chosen interval.

Advantages of isopleth map

1. It provides good visual impression if it is well presented
2. It is useful for showing distribution of phenomenon particularly climate.
- 3 The map preparation is fairly easy.
4. It can be analysed qualitatively

Disadvantage of isopleth map

- 1 .It is time consuming in preparation especially drawing
2. It is difficulty to quantify the data presented
- 3.It needs high skills to interpret data presented