Prof Dr Najlaa Fawzi
Family and Community Med Dept
2019-2020

ORDER ARRY: away of organizing data so that they are manageable.

Which is easier to understand?

3, 1, 7, 4, 1, 2, 3, 5, 4, 9

or

1, 1, 2, 3, 3, 4, 4, 5, 7, 9

PRESENTATION OF DATA:

Methods of presentation must be determined according to the data format, the method of analysis to be used, and the information to be emphasized

PRESENTATION OF DATA:

The aim of presenting data is to produce a precise and accurate demonstration of the information.

Summarized to simplify and highlighted to draw attention to the most important features.

This may be achieved by:

Text presentation

Mathematical: **a) Measures of central tendency**

b) Measures of dispersion.

Tabular: **using tables.**

Graphical: **using graphs**

Pictorial: **using pictures**

Text presentation

Text is the main method of conveying information as it is used to explain results and trends and provide contextual information.
Data are fundamentally presented in paragraphs or sentences.
Text can be used to provide interpretation or emphasize certain data.
If quantitative information to be carried consists of one or two numbers, it is more appropriate to use written language than tables or graphs.

For example, information about the incidence rates of delirium following anesthesia in 2016–2017 can be presented with the use of a few numbers: “The incidence rate of delirium following anesthesia was 11% in 2016 and 15% in 2017; no significant difference of incidence rates was found between the two years.”

If this information were to be presented in a graph or a table, it would occupy an unnecessarily large space on the page, without enhancing the readers' understanding of the data.

If more data are to be presented, or other information such as that regarding data trends are to be conveyed, a table or a graph would be more appropriate.

By nature, data take longer to read when presented as texts and when the main text includes a long list of information, readers and reviewers may have difficulties in understanding the information.

Data Presentation

- Principals of data presentation

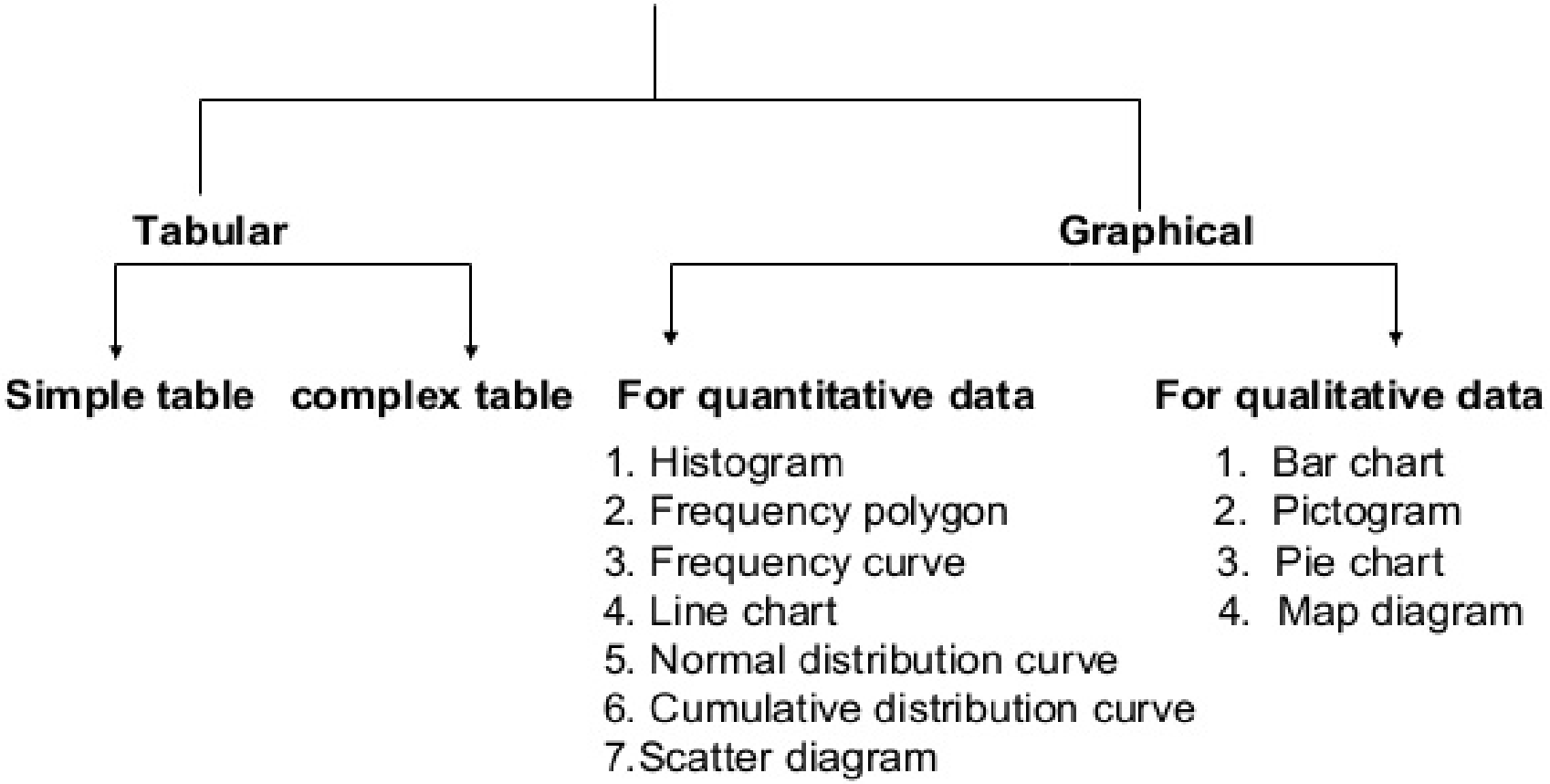
(a) To arrange the data in such a way that it should create interest in the reader's mind at the first sight.

(b) To present the information in a compact and concise form without losing important details.

(c) To present the data in a simple form so as to draw the conclusion directly by viewing at the data.

(d) To present it in such away that it can help in further statistical analysis.

Presentation of data



TABULAR PRESENTATION OF DATA:

FREQUENCY DISTRIBUTION:

A frequency table is used to summarize categorical or numerical data.

TABULAR PRESENTATION OF DATA:

1.Single variable frequencies: Frequency distribution table .

a)For qualitative variables: we must put them into groups & with in each group we put frequencies.

Table-1 Frequency distribution of peptic ulcer according to site of ulcer

Site of ulcer	Frequency	Percent %
Gastric ulcer	24	30
Duodenal ulcer	50	62.5
Gastric and duodenal ulcer	6	7.5
Total	80	100

The main task being to count the number of observation in each category , these counts are called the frequency .

They are often presented as relative frequency that is the percentage of the total number of individuals.

b-For quantitative variables , for discrete variables the frequencies may be tabulated either for each value of the variable or for groups of values .

History of abortion of 50 women

abortion	frequency	RF	RF%	CUM frq	Cum RF	Cum RF%
0	2	0.04	4	2	0.04	4
1	5	0.10	10	7	0.14	14
2	8	0.16	16	15	0.3	30
3	10	0.20	20	25	0.5	50
4	15	0.30	30	40	0.8	80
5	7	0.14	14	47	0.94	94
6	3	0.06	6	50	1	100
total	50	1	100			

Frequency Distribution for Discrete Random Variables

Example:
Suppose that we take a sample of size 16 from children in a primary school and get the following data about the number of their decayed teeth,
3,5,2,4,0,1,3,5,2,3,2,3,3,2,4,1
To construct a frequency table:
1- Order the values from the smallest to the largest.
0,1,1,2,2,2,2,3,3,3,3,3,4,4,5,5
2- Count how many numbers are the same

No. of decayed teeth	Frequency	Relative Frequency
0	1	0.0625
1	2	0.125
2	4	0.25
3	5	0.3125
4	2	0.125
5	2	0.125
Total	16	1

C- For a large data set on a quantitative variable requiring grouping of the data into classes.

CLASS INTERVALS: a set of contiguous, non-overlapping intervals, it is used to group a set of observations such a way that each value can be placed on one interval only.

STURGE'S RULE: used to decide the number & width of class intervals:

$$K = 1 + 3.322 \log n \qquad \& \qquad W = R / K$$

Where K = no. of intervals

N = no. of observations = total no. of measurements

W = width of internals

R = the range of readings = largest value (L) – smallest value (S)

FREQUENCY DISTRIBUTION (F.D): number of individuals falling into each class of interval.

While Relative Frequency Distribution: is the proportion of values in each class interval, which is determined by dividing F.D by total no. of observations. Cumulative Frequency Distribution & Cumulative Relative Frequency Distribution are used to facilitate obtaining information.

Example: weights of malignant tumors removed from the abdomen of 57 subjects:

68, 63, 42, 27, 30, 36, 28, 32, 79, 27, 22, 23, 24, 25, 44, 65, 43, 25, 74, 51, 36, 42, 28, 31, 28, 25, 45, 12, 57, 51, 12, 32, 49, 38, 42, 27, 31, 50, 38, 21, 16, 24, 69, 47, 23, 22, 43, 27, 49, 28, 23, 19, 46, 30, 43, 49, 12.

$K = 1 + 3.322 \log (57) = 1 + 3.322 (1.7559) = 6.833 \approx 7$

$W = R / K = (79 - 12) / 7 = 9.6 \approx 10$

Class interval	f	cf	R.F. %	C.R.F. %
10 - 19	5	5	8.77	8.77
20 - 29	19	24	33.33	42.1
30 – 39	10	34	17.54	59.64
40 – 49	13	47	22.81	82.45
50 – 59	4	51	7.02	89.47
60 – 69	4	55	7.02	96.49
70 – 79	2	57	3.51	100.00
Totals	57	---	100.00	-------

Class interval	Mid – interval	Frequency Freq (f)	Cumulative Frequency	Relative Frequency R.f	Cumulative Relative Frequency
30 – 39	34.5	11	11	0.0582	0.0582
40 – 49	44.5	46	57	0.2434	-
50 – 59	54.5	-	127	-	0.6720
60 – 69	-	45	-	0.2381	0.9101
70 – 79	74.5	16	188	0.0847	0.9948
80 – 89	84.5	1	189	0.0053	1
Total		189		1	

2-compound table

A compound table is just an extension of a simple in which there are more than one variable distributed among its attributes (sub-variable).

This may refer to a compound table as a cross tabulation or even to a contingency table depending on the context in which it is used.

Cross Tabulation:

is a method to quantitatively analyze the relationship between multiple variables. Also known as contingency tables
or cross tabs, cross tabulation groups variables to understand the correlation between different variables

Is a joint frequency distribution of different values of two (or more) variables.

A-Two dimensional tables.

B-Three dimensional tables. As in age, gender & smoking

Table 2: Frequency distribution of adult patients by Hb and gender:

Hb (g/dl)	Gender		Total
	Male	Female	
<9.0	0	2	2
9.0 – 9.9	1	3	4
10.0 – 10.9	3	5	8
11.0 – 11.9	6	8	14
12.0 – 12.9	10	6	16
13.0 – 13.9	5	4	9
14.0 – 14.9	3	2	5
15.0 – 15.9	2	0	2
Total	30	30	60

There are advantages in such table presentation. One advantage is that more than one parameter (education and occupation) can be shown in the same table. This serves as a brief presentation. Another advantage is easy comparison and interpretation.

Principles of making tables:

1. **Simple**: the table must be simple; it should contain least no. of variables to make things easy to the reader.
It is better to have 2-3 tables than to have one complicated table.

2. **Understandable & self explanatory**: we should make the table easy to be understand without the need to return back to the text, & this is done by:

a) Using symbols, codes & abbreviations which must be explained by putting down a (foot note) at the bottom of the table.

b-Clear concise labeling of rows & columns.

C-The units of data must be defined, e.g. when we want to measure the height we must put cm or meter.

d-Using clear concise title & the question what where & when should be answered within the title, placed at top

table for infant mortality rate for 2008in Baghdad.

e. The total should be placed as in frequency distribution; we give the total at the end.

Types of tables

Simple tables

Compound tables: consists of 2 or more rows & columns.

Contingency tables: relationship between one or more variables

Table7: The distribution of study group according to place of residence and family type with inter pregnancy intervals

Total (n=500)				Inter pregnancy interval						χ^2;d.f.;P value
				<2Y (n=205)		2-4 (n=226)		>4Y (n=69)		
		No	%	No	%	No	%	No	%	
Place of residency	Rural	156	31.2	64	41	71	45.5	21	13.5	0.029;2; 0.986
	Urban	344	68.8	141	41.0	155	45	48	14.0	
Family type	Nuclear	188	37.6	67	35.6	91	48.4	30	16.0	3.813;2; 0.149
	Extended	312	62.4	138	44.2	135	43.3	39	12.5	

Graphic Presentation

Graphical presentation of data has also become quite popular.
It gives visual information in addition to magnitudes.

Furthermore, comparisons and changes in the data can be well visualized when presented in graphical form.
A very useful part of graphical presentation is the interpretation of the graphs.
In every graph should try to understand the data.

Importance of Diagrammatic Representation

1-They have greater attraction than simple figures.

They give delight to the eye and add a spark of interest.

2-They help in deriving the required information in less time and without any mental strain.

3-They facilitate comparison.

4-They may reveal unsuspected patterns in a complex set of data and may suggest directions in which changes are occurring.

5-They have greater memorising value than mere figures.

This is so because the impression left by the diagram is of a lasting nature.

GRAPHICAL PRESENTATION OF DATA

pictorial display of data using coordinate measures.

X-axis is for independent variable (method of classification), & Y-axis is for dependant variable (frequency or relative frequency).

General principles to make a graph:

Simple: doesn't contain so much lines & symbols more than the eye can follow (more easier to the reader to understand the graph) self explanatory

Title: placed on the bottom of the graph & answer the questions of what, where & when.

Keys: if there is more than one variable, we must notify what each variable represent.

scale & units: we must write the units which are used in the graph

Types of graphs:

There are a wide variety of ways of presenting data pictorially such as : Bar chart; Pie chart; Histogram ; Frequency polygon; Cumulative frequency curve; Scatter diagram; Flowchart; Geography coordinate charts; Maps.

Bar chart: is used for variables of qualitative data.

Is a tool for comparing categories of mutually exclusive data. Both nominal and ordinal data

The different categories usually are indicated on the x- axis, the frequency of data of each category is indicated on the Y- axis, and the categories are compared by the height of the bars, since the data categories discrete , there is a gap between the bars, and the bars can be arranged in any order on the x- axis with space between them.

It is important that Y- axis should start at zero.

Bar chart differs from histogram: used for qualitative (discrete) variables & can be used for more than 1 set of data.

Simple bar charts have two purposes.

1) They reflect the actual magnitude of the frequency of each item.

2) Their frequencies can be compared by comparing the heights of bars on the chart.

Component bar charts are used to sum each component.

Multiple bar charts

Component bar charts and multiple bar charts are similar, except for the following:

1. Component bar charts can show the grand total, whereas multiple bar charts cannot.

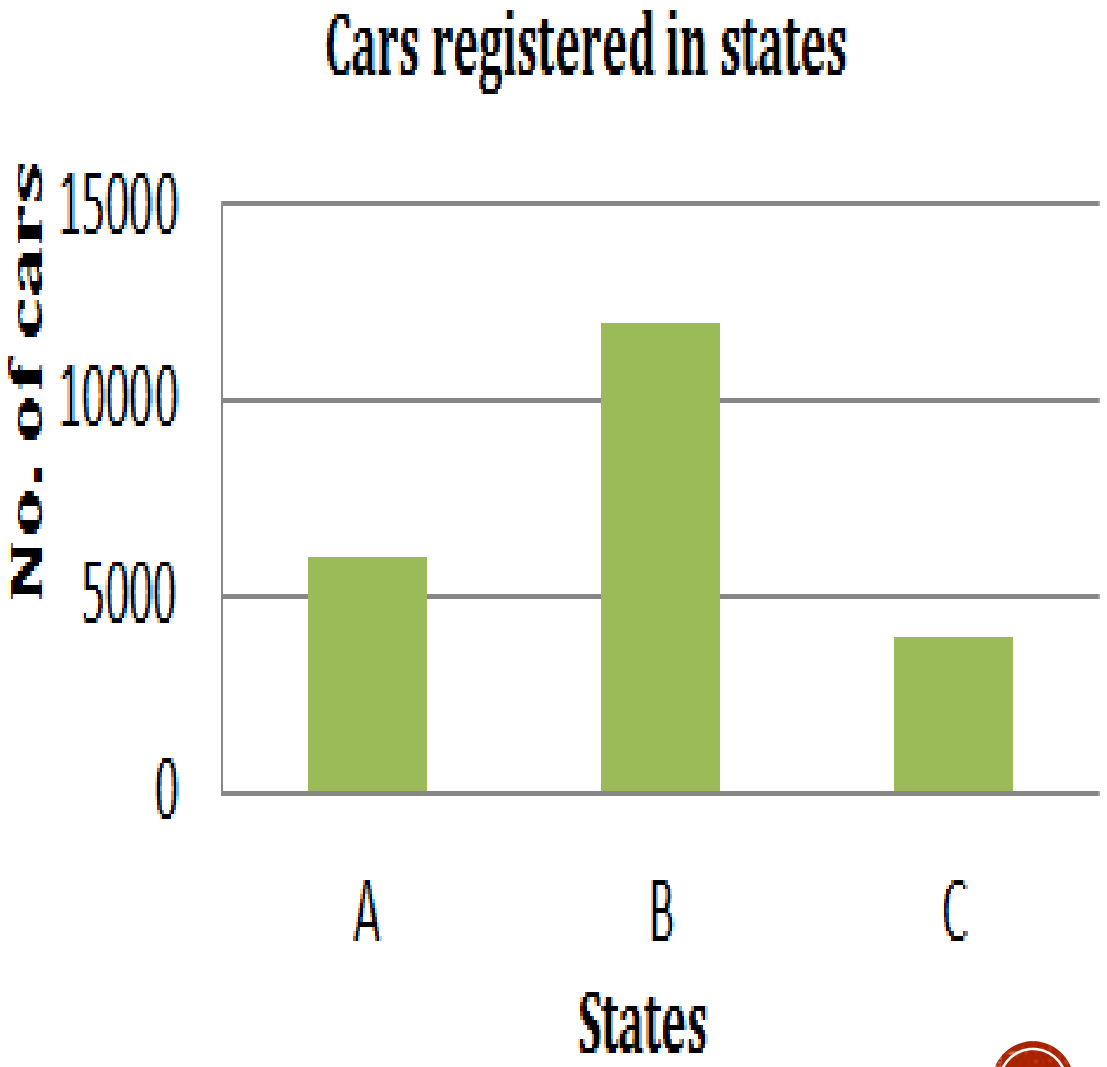
2. Component bar charts cannot show clear comparative magnitudes of each component, whereas multiple bar charts can.

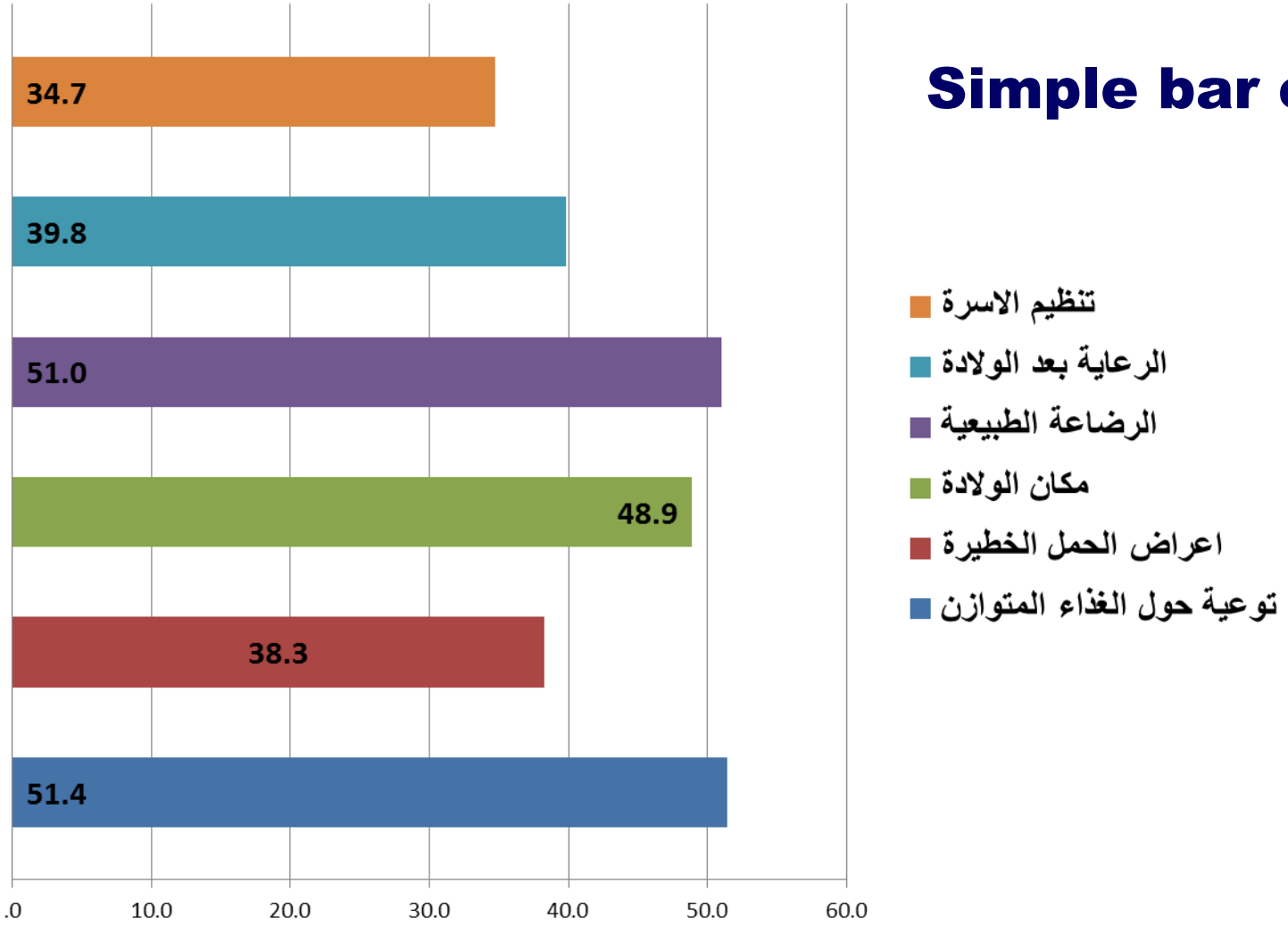
Single Bar Chart

Number of Cars registered in 3 States

States	No. of Cars
A	6000
B	12,000
C	4,000

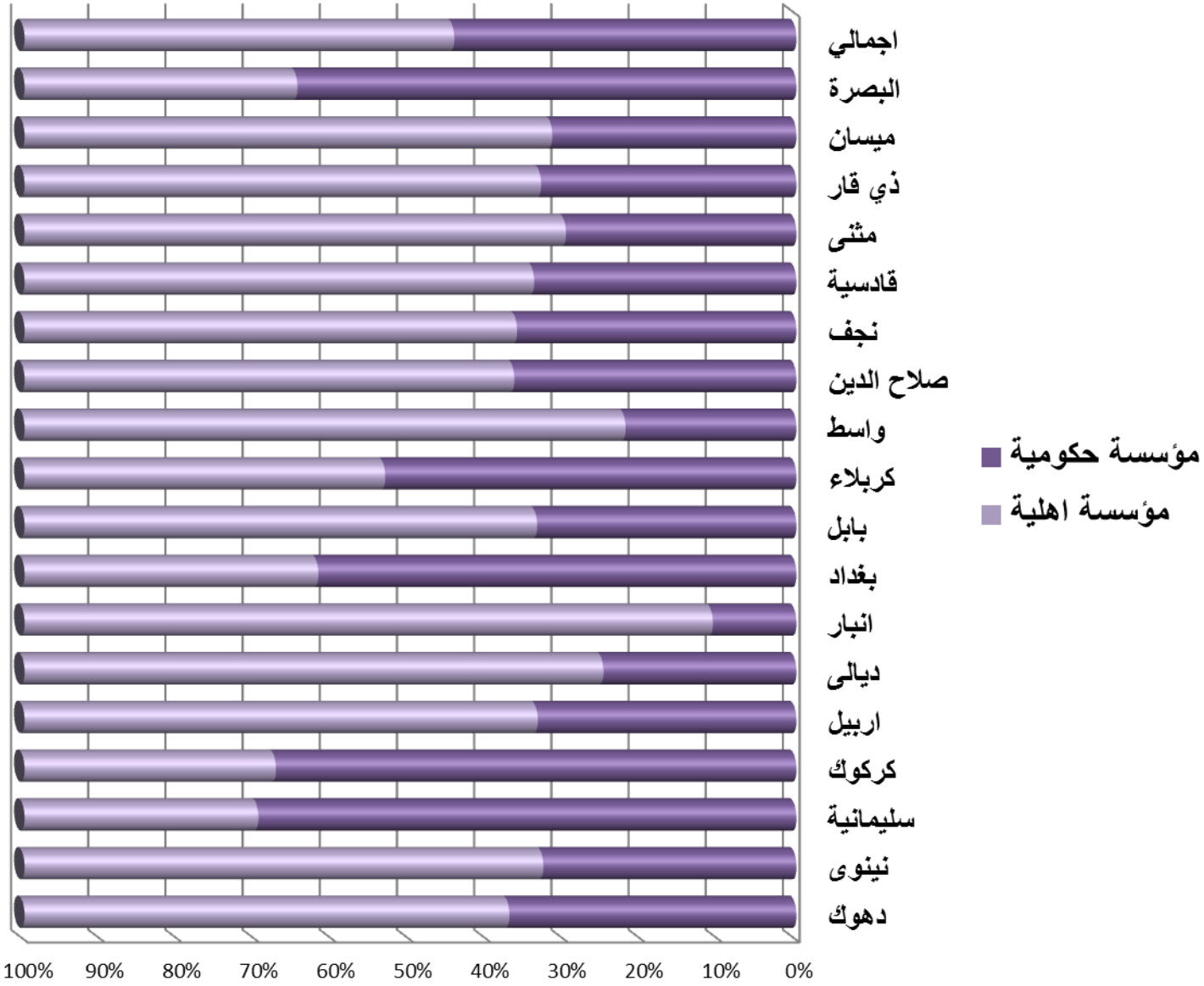
Simple bar chart





الشكل رقم-6 التوزيع النسبي للمعلومات التي حصلت عليها النساء خلال الرعاية خلال الحمل –العراق-2011

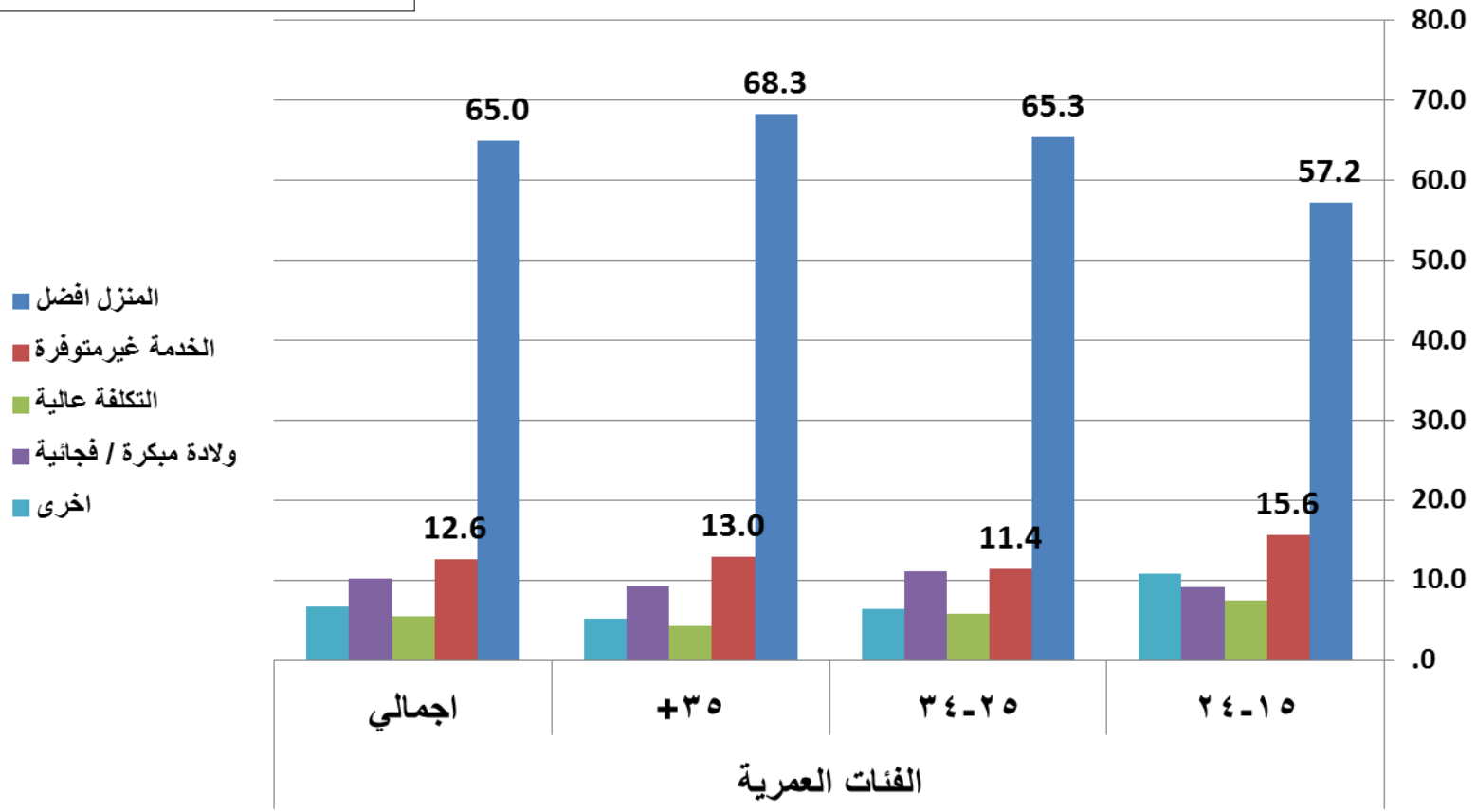
Component bar chart



MULTIPLE BAR CHART

A chart depicting two or more characteristics in the form of bars of length proportional in magnitude of the characteristics.
For example, a chart comparing the age and sex distribution of two populations may be drawn with sets of bars, one bar of each pair for each population, and one pair for each age group.

multiple bar chart



الشكل رقم-12 التوزيع النسبي للنساء حسب الفئات العمرية واسباب الولادة في المنزل-العراق-2011

Pie chart: it is a circle divided into sectors with areas proportional to the frequencies or the relative frequencies of the categories of the variable.

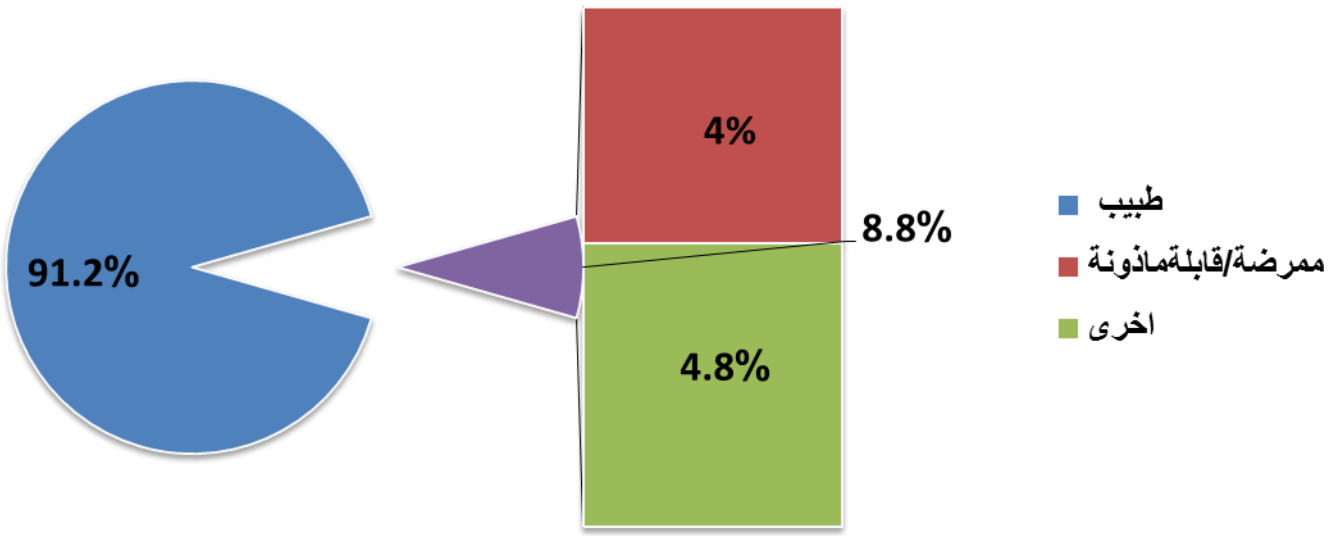
It is used for one set of data. It represents the part from the total; each portion of the circle illustrates the division of the whole into segments.

The best way to start to draw a pie chart is to start at 12 o'clock & with clockwise direction beginning with largest to lowest frequency.

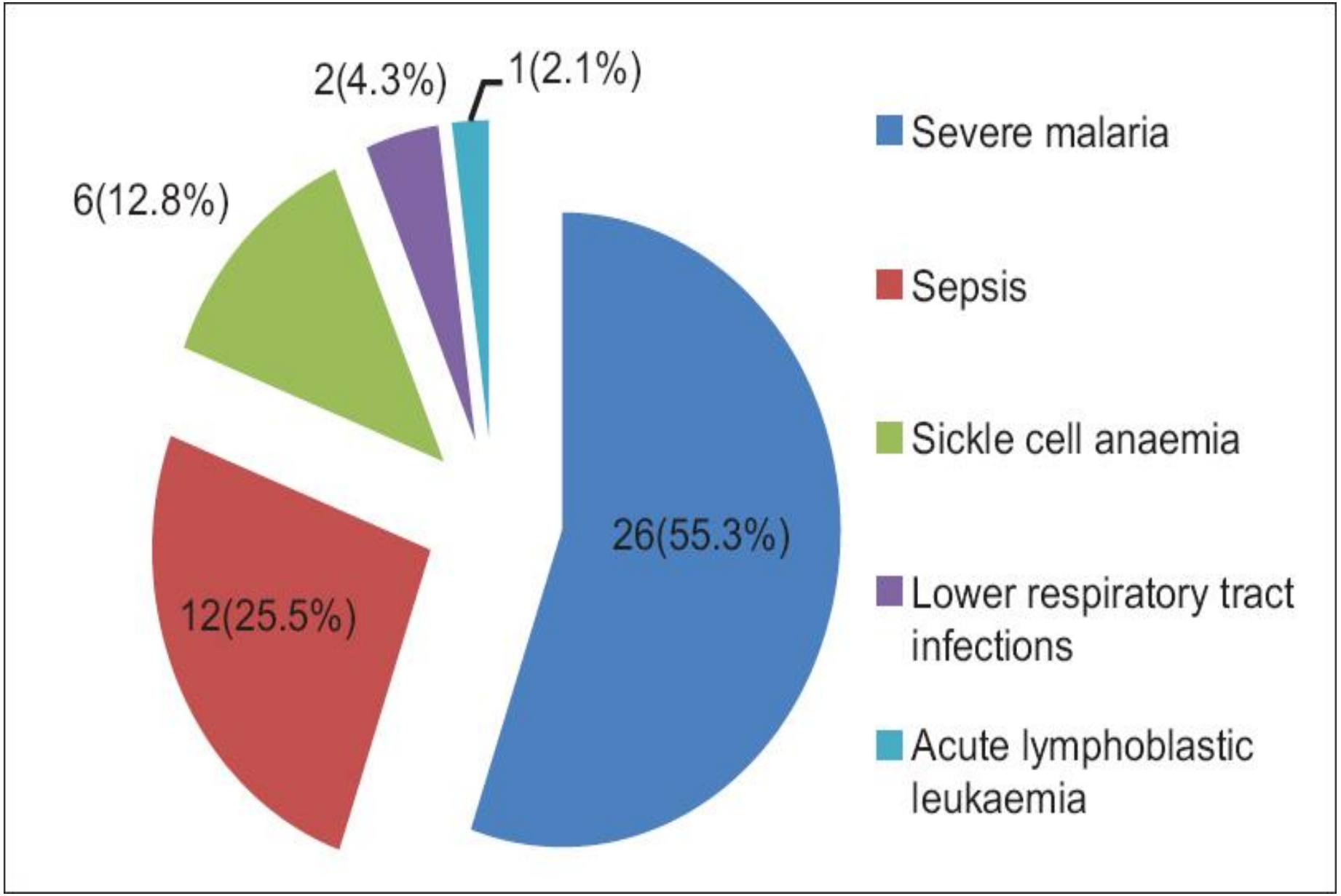
To represent the data as pie chart we must :

- Find the relative frequency distribution of each category (i.e. % of each variable).

- Multiply the relative frequency distribution by 360 to find the degree of each category.



الشكل رقم-21 التوزيع النسبي للنساء اللاتي تلقين رعاية لما بعد الولادة حسب مقدم الرعاية- العراق2011


2(4.3%)
1(2.1%)
6(12.8%)
26(55.3%)
12(25.5%)
Severe malaria
Sepsis
Sickle cell anaemia
Lower respiratory tract infections
Acute lymphoblastic leukaemia

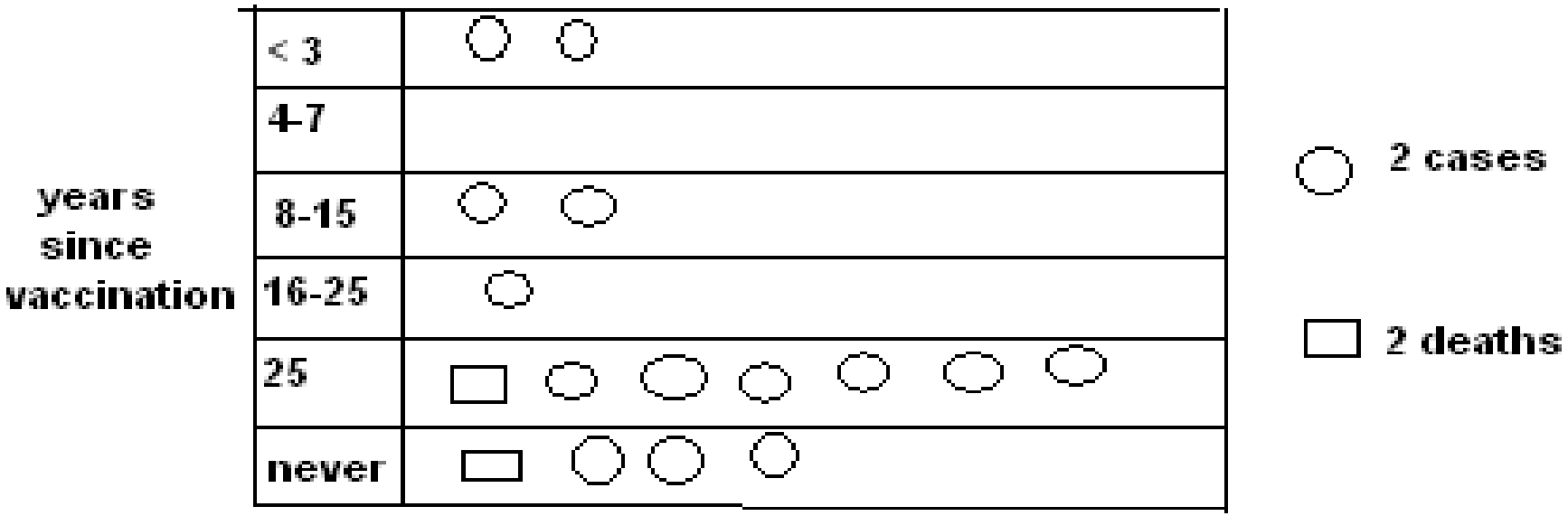
Pictogram/Picture Diagram

— Is a method to impress the frequency of occurrence of events to common man

A pictogram or pictograph represents the frequency of data as pictures or symbols. Each picture or symbol may represent one or more units of the data.

Pictogram: it uses a series of small identifying symbols to present the data, each symbol represent a fixed no. of items .

Number of pictures give quick idea of the relative importance of the problem.



Frequency Graphs : The graphical representation of a frequency distribution is known as a frequency graph.

There are four representation : 1- line diagram 2- histogram 3- frequency polygon 4 – cumulative frequency curve

The line diagram

The line graph is especially useful for the study of some variables according to the passage of time. The time, in weeks, months or years is marked along the horizontal axis; and the value of the quantity that is being studied is marked on the vertical axis.
The distance of each plotted point above the base-line indicates its numerical value.

The line graph is suitable for depicting a consecutive trend of a series over a long period.

line graph

Monthly reported Rubella cases,USA,1980

X-axis represents independent variable (months) while Y-axis represents dependant variables (cases of Rubella per 105).

Line diagram is used to show the trend of disease or event over time.

Line diagram is also used to Present discrete quantitative data .

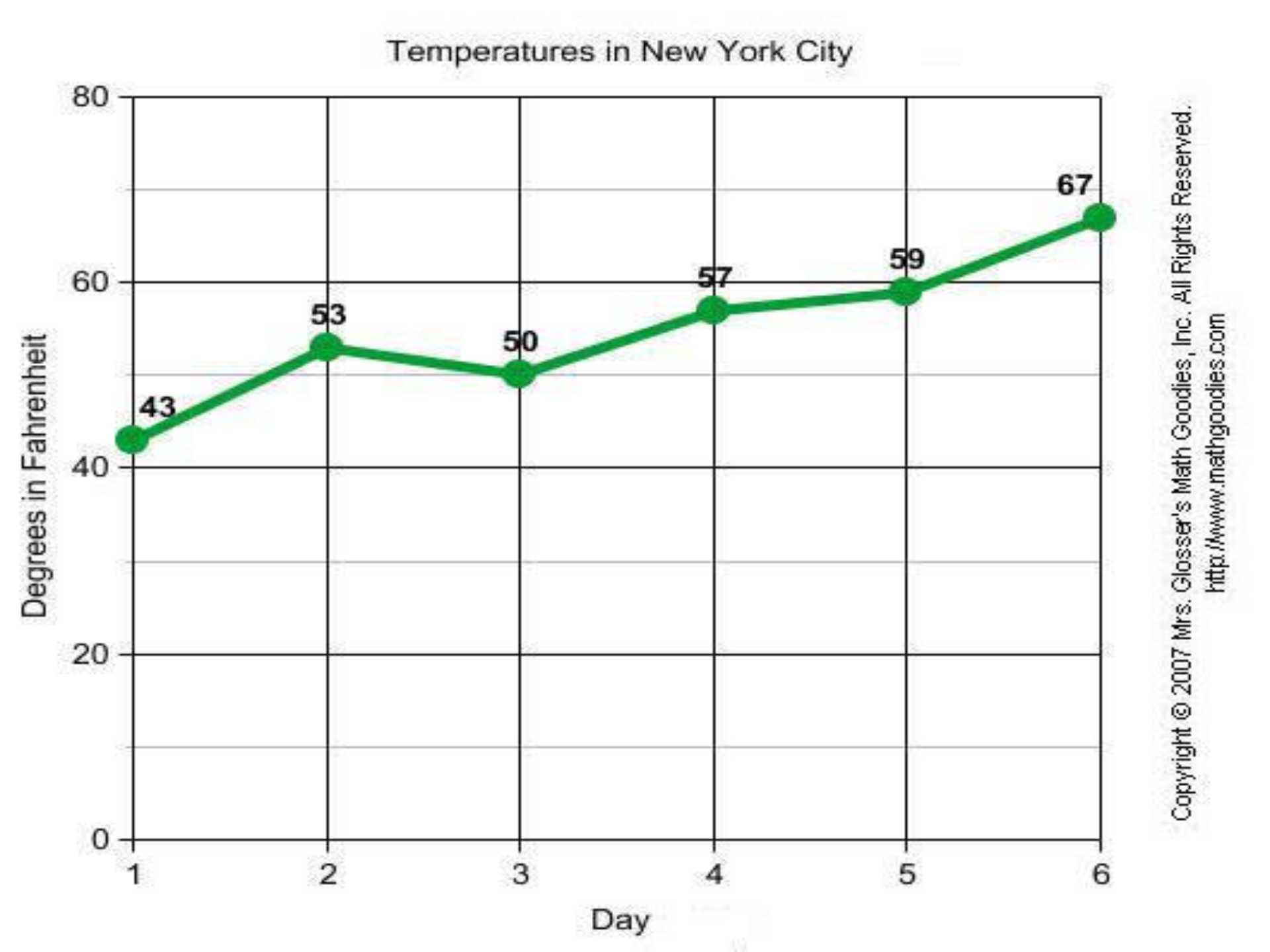
Line diagrams are drawn by plotting the values of two continuous variables.

These show trends or changes in one variable resulting from changes in the other.

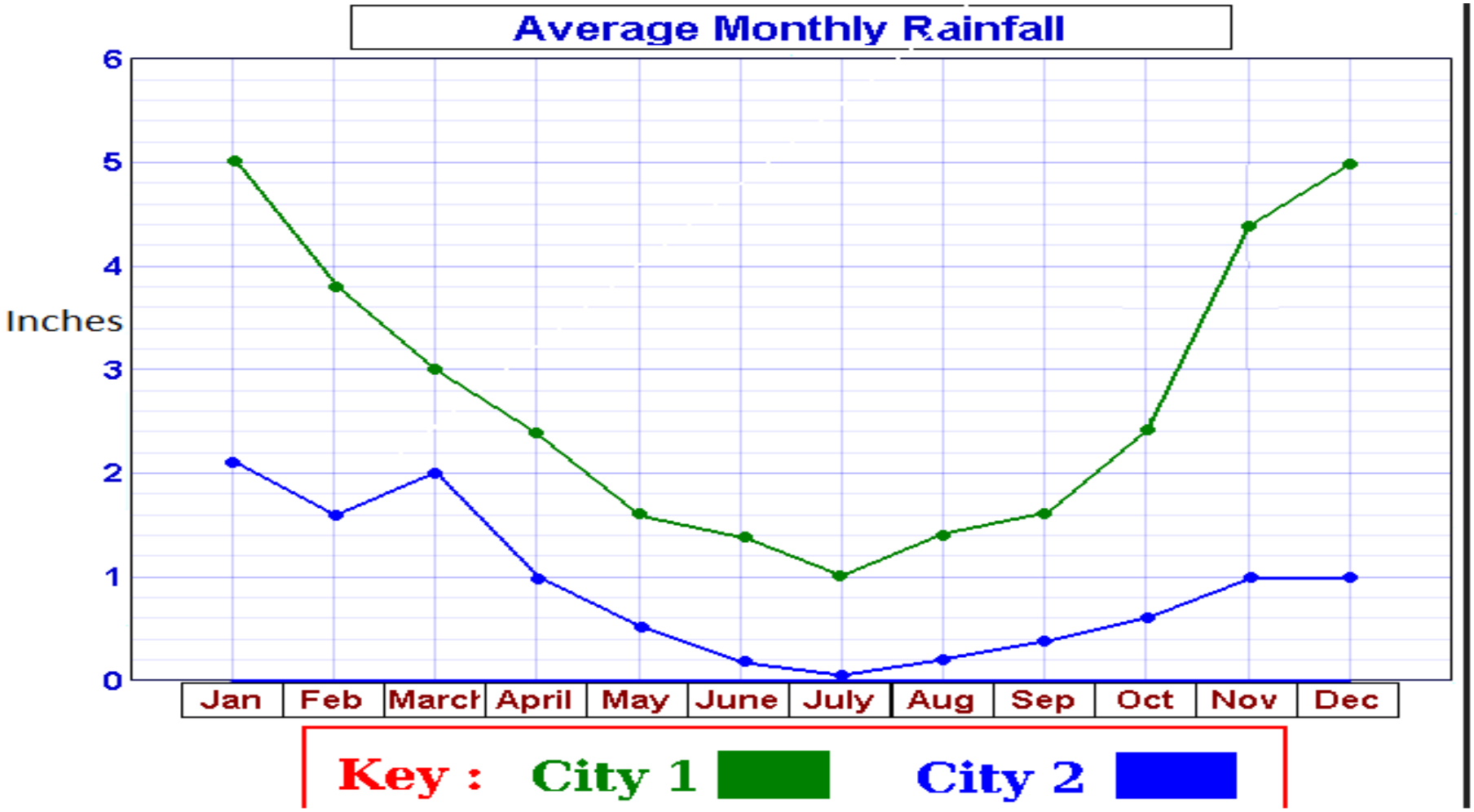
One important application of the line diagram is to study the changes of various economic indicators over time.

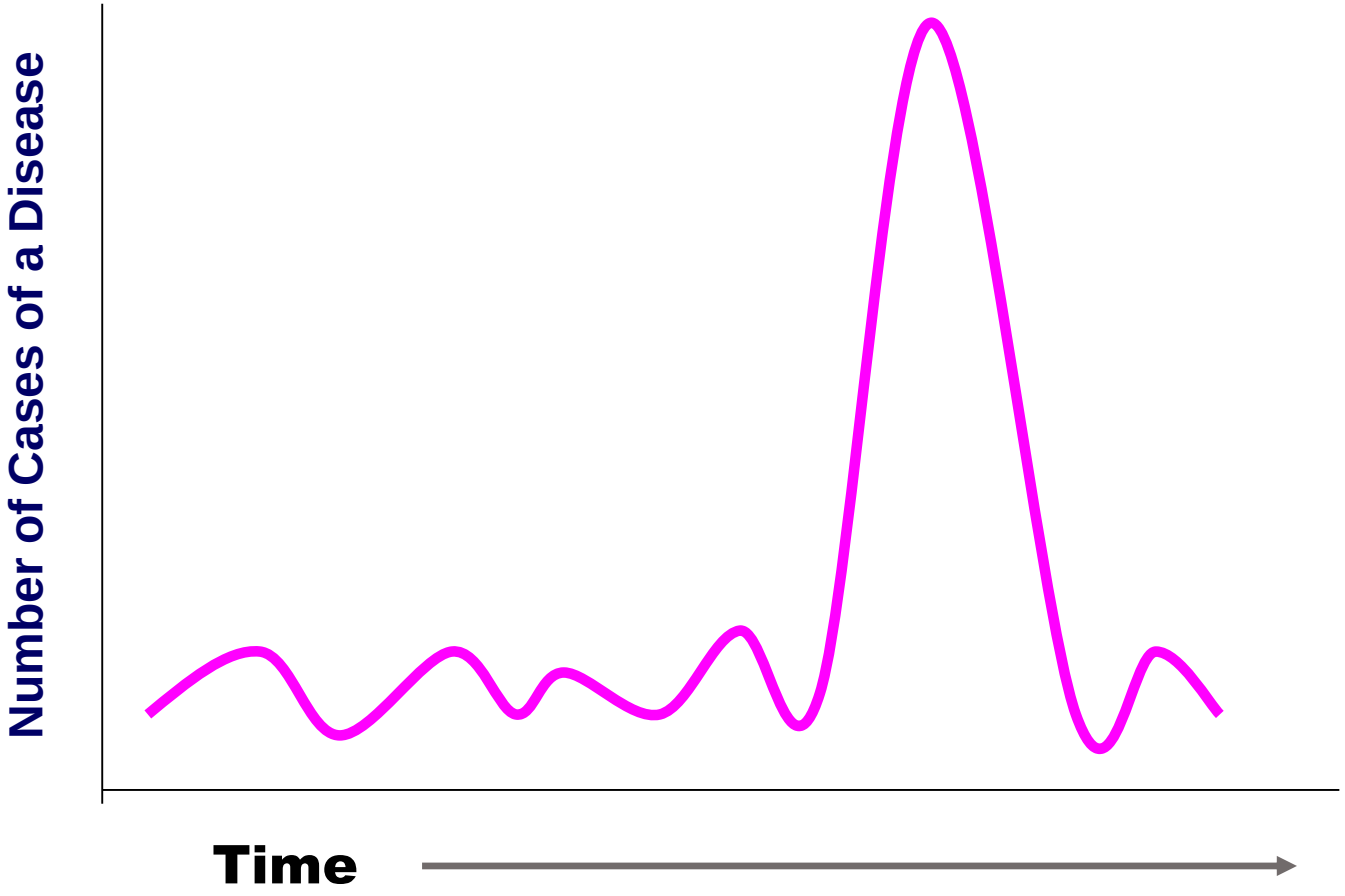
Line diagrams may be presented in the form of continuous lines or segmented lines depending on the phenomenon under study.

Single line graph The simplest of line graphs is the single line graph, so called because it displays information concerning one variable only, in terms of its frequencies.

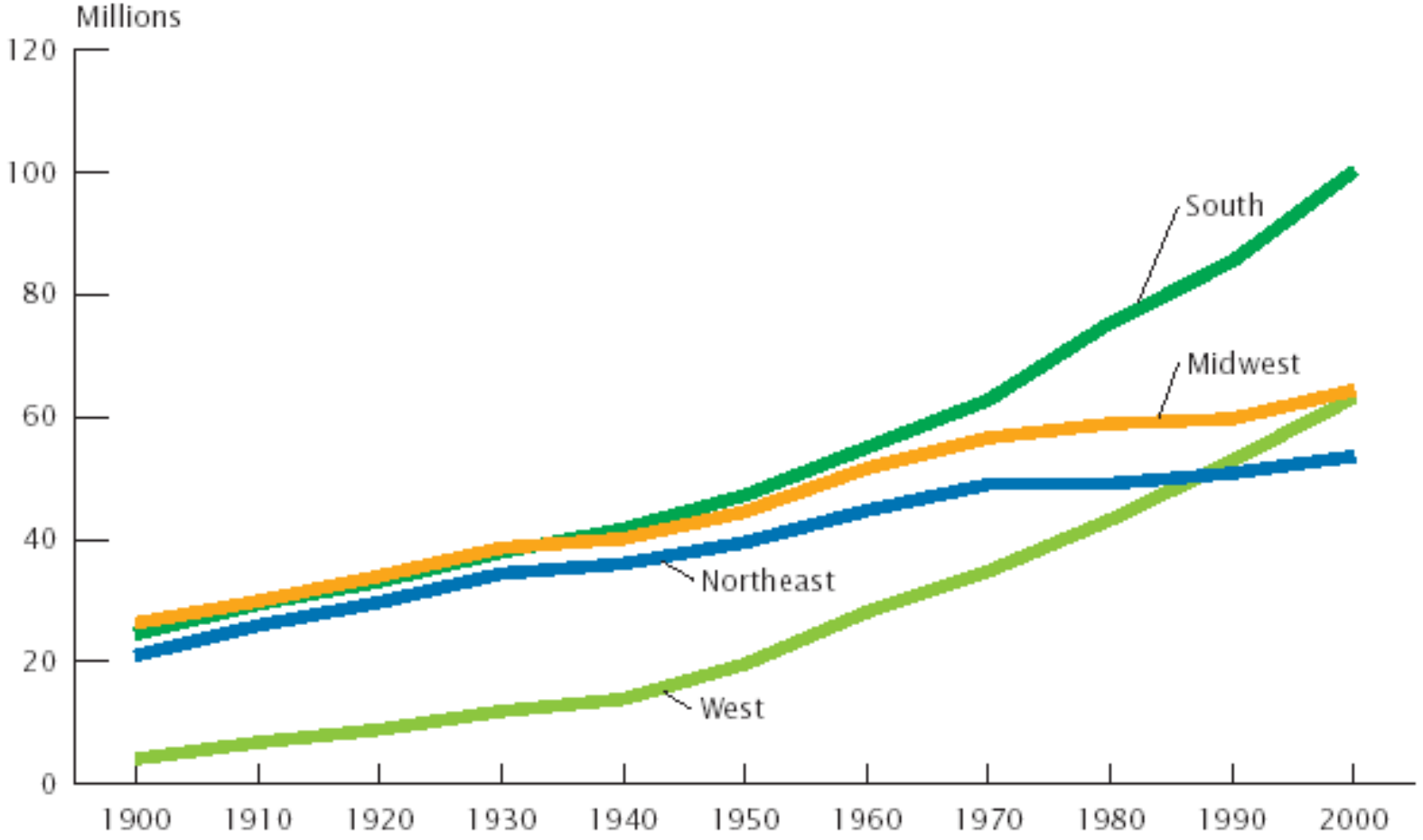

Temperatures in New York City
Degrees in Fahrenheit
80
60
40
20
0
1
2
3
4
5
6
Day
43
53
50
57
59
67
Copyright © 2007 Mrs. Glosser's Math Goodies, Inc. All Rights Reserved.
http://www.mathgoodies.com

Multiple line graphs illustrate information on several variables so that comparison is possible between them



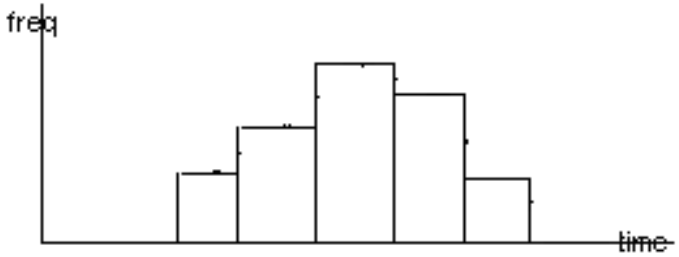

Number of Cases of a Disease
Time

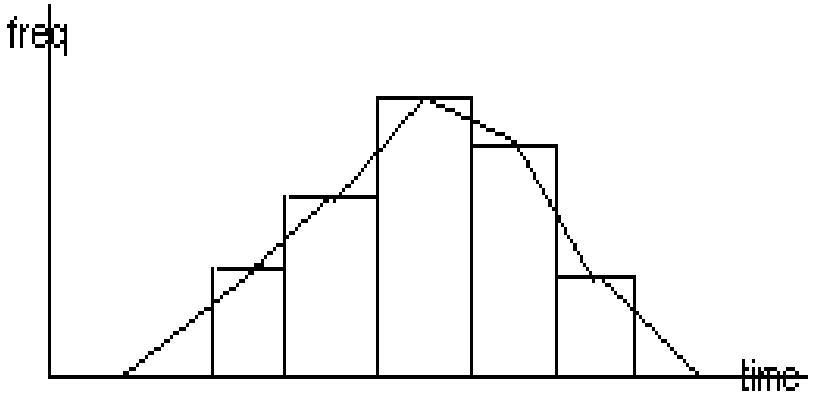
Total Population by Region: 1900 to 2000



Source: U.S. Census Bureau, decennial census of population, 1900 to 2000.

Histogram





Histogram: used for quantitative variable & for one set of data, if we have a histogram which has certain variables concerned with males, we must make another histogram for females (if we want to make another histogram for females, but we cannot mix both genders of the same histogram.

Frequency histogram is used to represent the class intervals so the frequency distribution will be represented as Y-axis while values of the variable

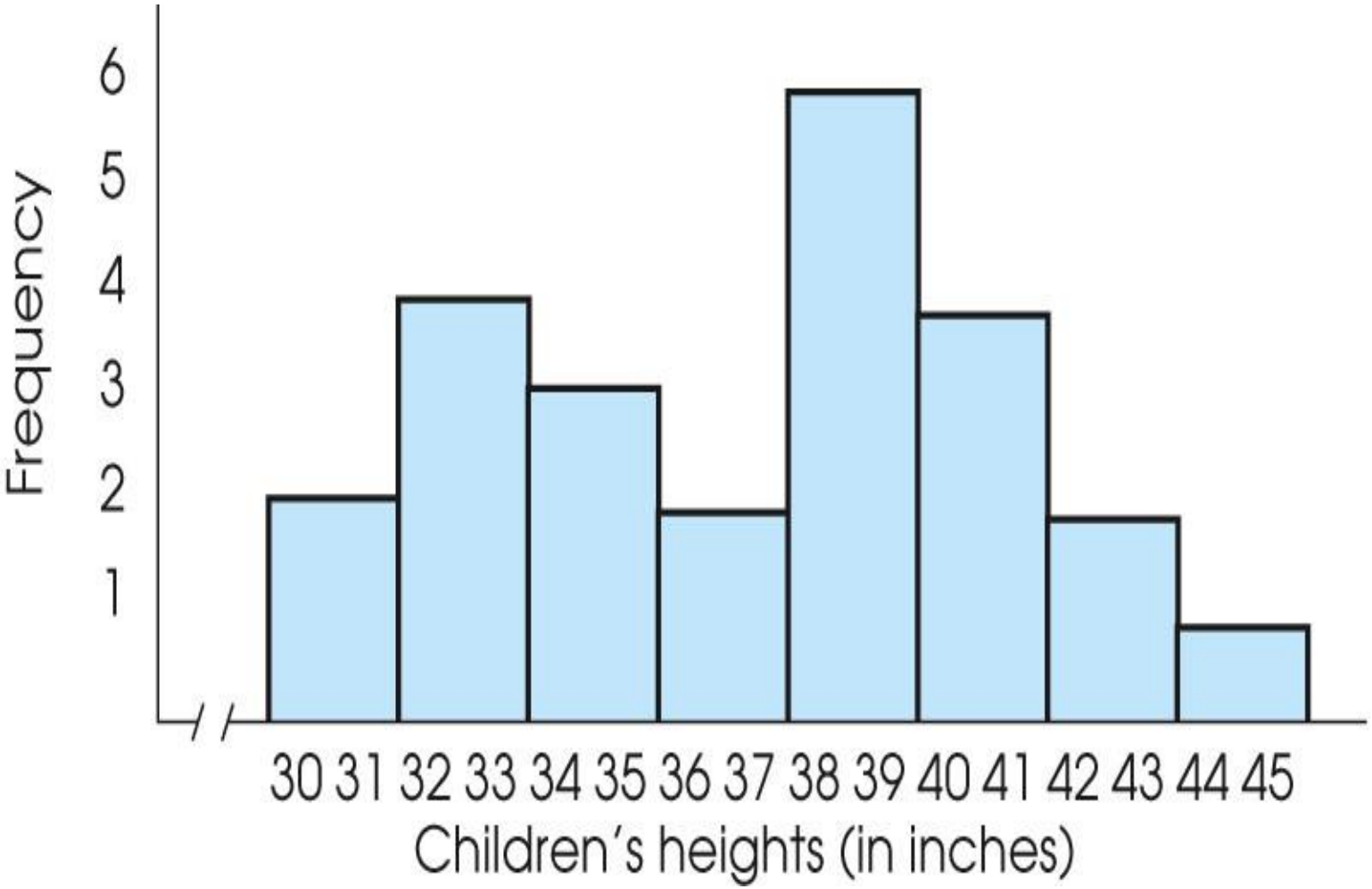
(i.e. categories of class interval) will be represented as X axis,

so that net result is no. of rectangles with width which represent the width of each category in class interval & height which represent the frequency distribution of that category.

Width of rectangles is the same as it equal to the width of class interval.

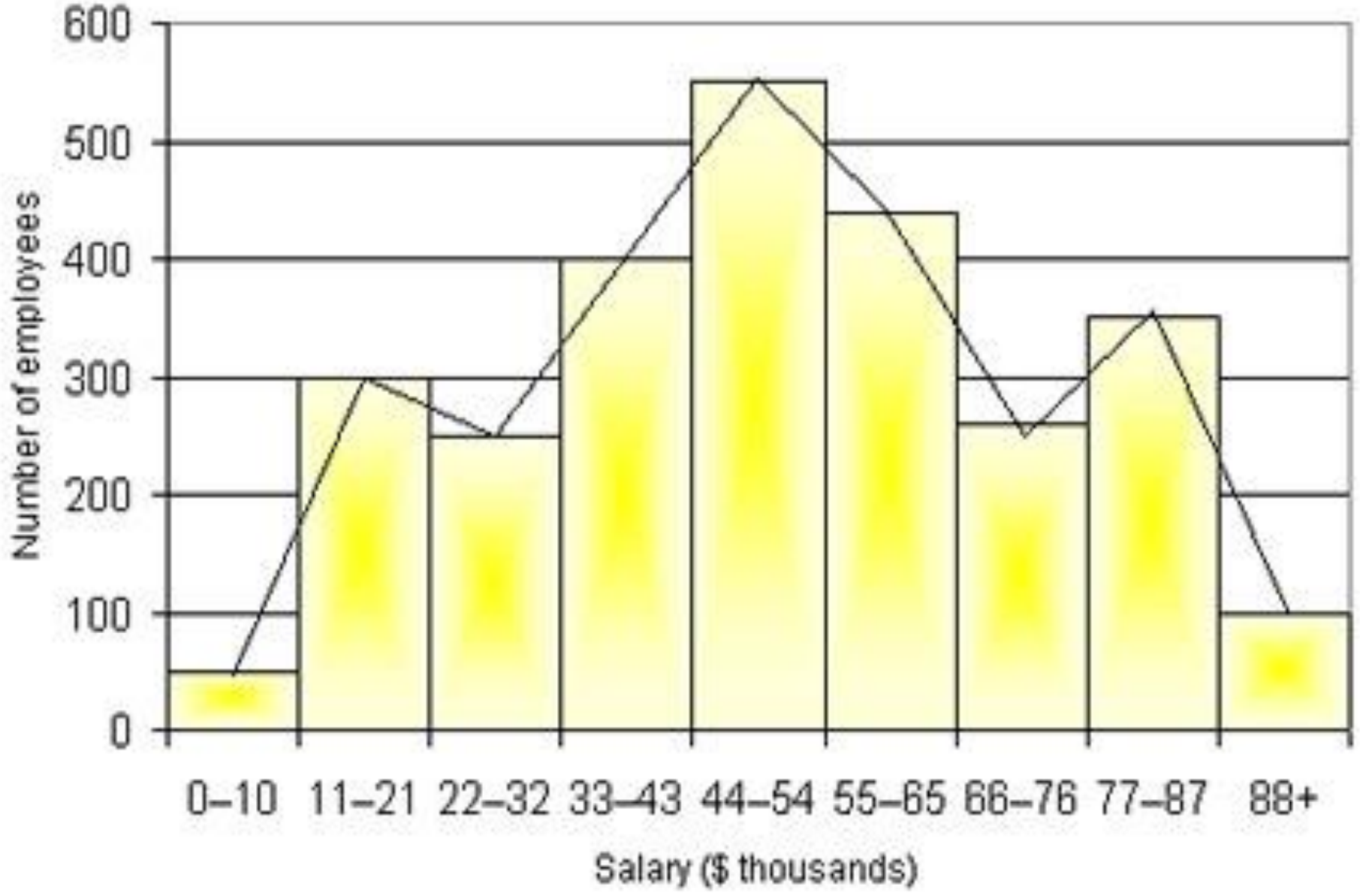
Rectangles are not separated from each other as the data in the class interval is continuous.

If data not start from zero in any axis, this axis must be marked for that (with a small zigzag).

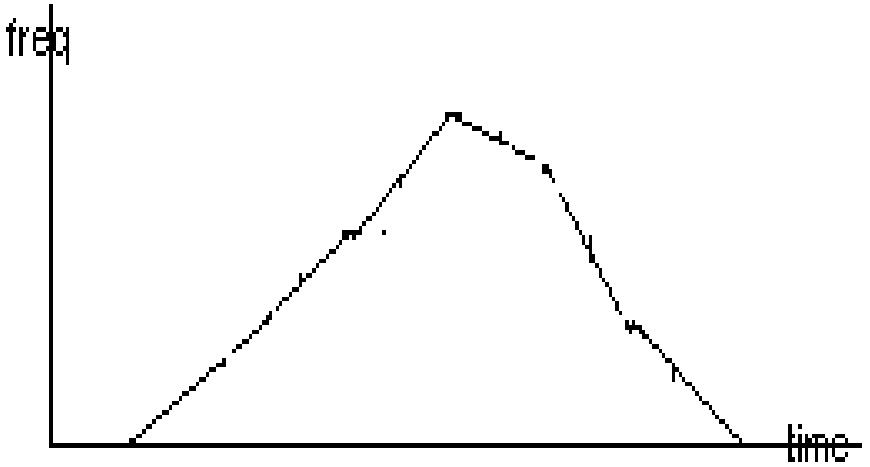


X	f
44–45	1
42–43	2
40–41	4
38–39	6
36–37	2
34–35	3
32–33	4
30–31	2

Figure 2. Distribution of salaries for the Acme Corporation



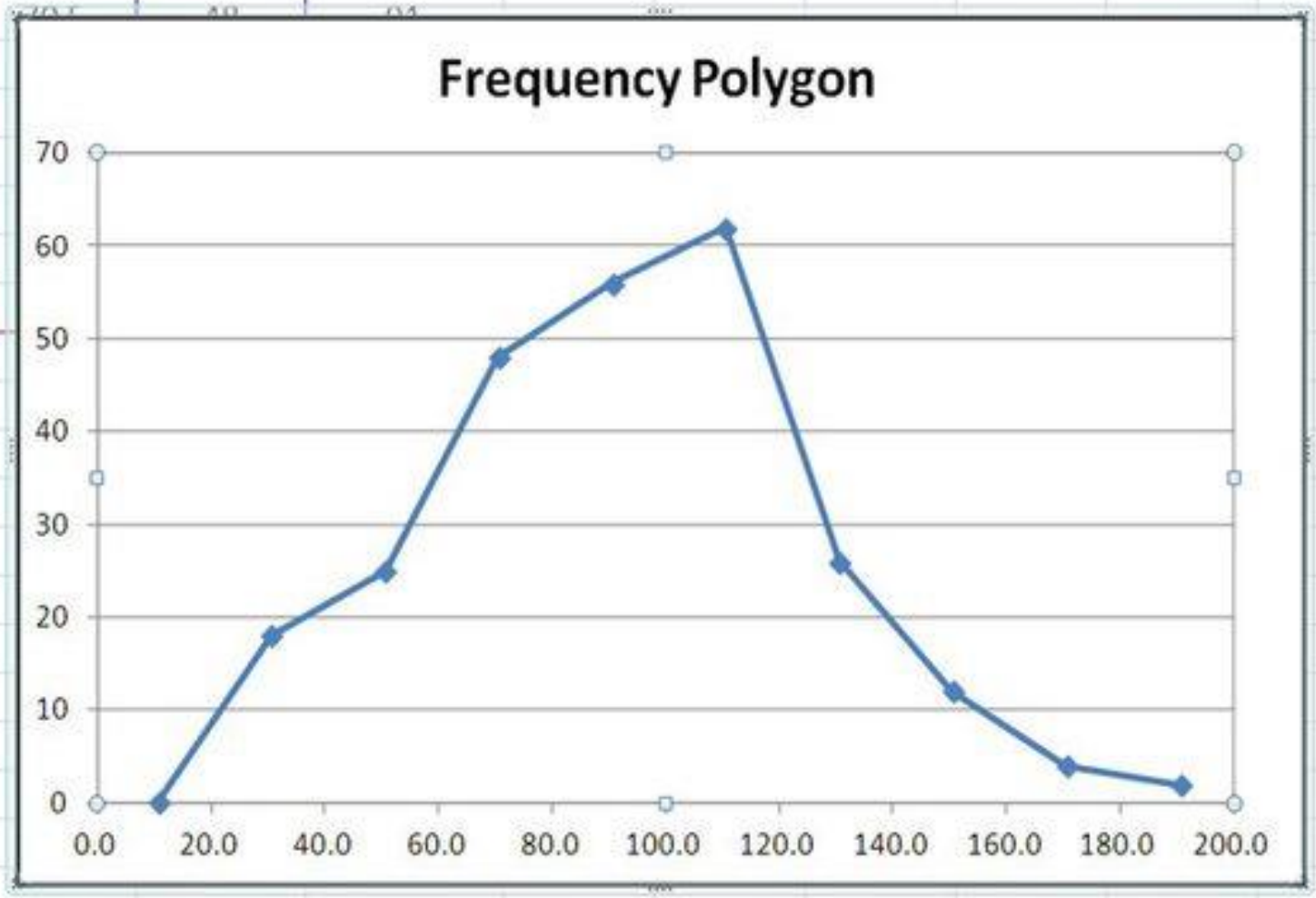
Frequency polygon



Frequency polygon: it is similar to histogram in its use for quantitative variable but polygon can be used for 2 or more sets of data & this is an advantage of this polygon in facilitating comparisons.

Another advantage of the polygon that its curvature give an idea about the distribution of data (normally distributed or not).

	C	D	E	F
1	Classes	Mid-Point	Frequency	Cumulative Frequency
2	1 - 20	10.5	0	0
3	21 - 40	30.5	18	18
4	41 - 60	50.5	25	43
5	61 - 80			
6	81 - 100			
7	101 - 120			
8	121 - 140			
9	141 - 160			
10	161 - 180			
11	181 - 200			

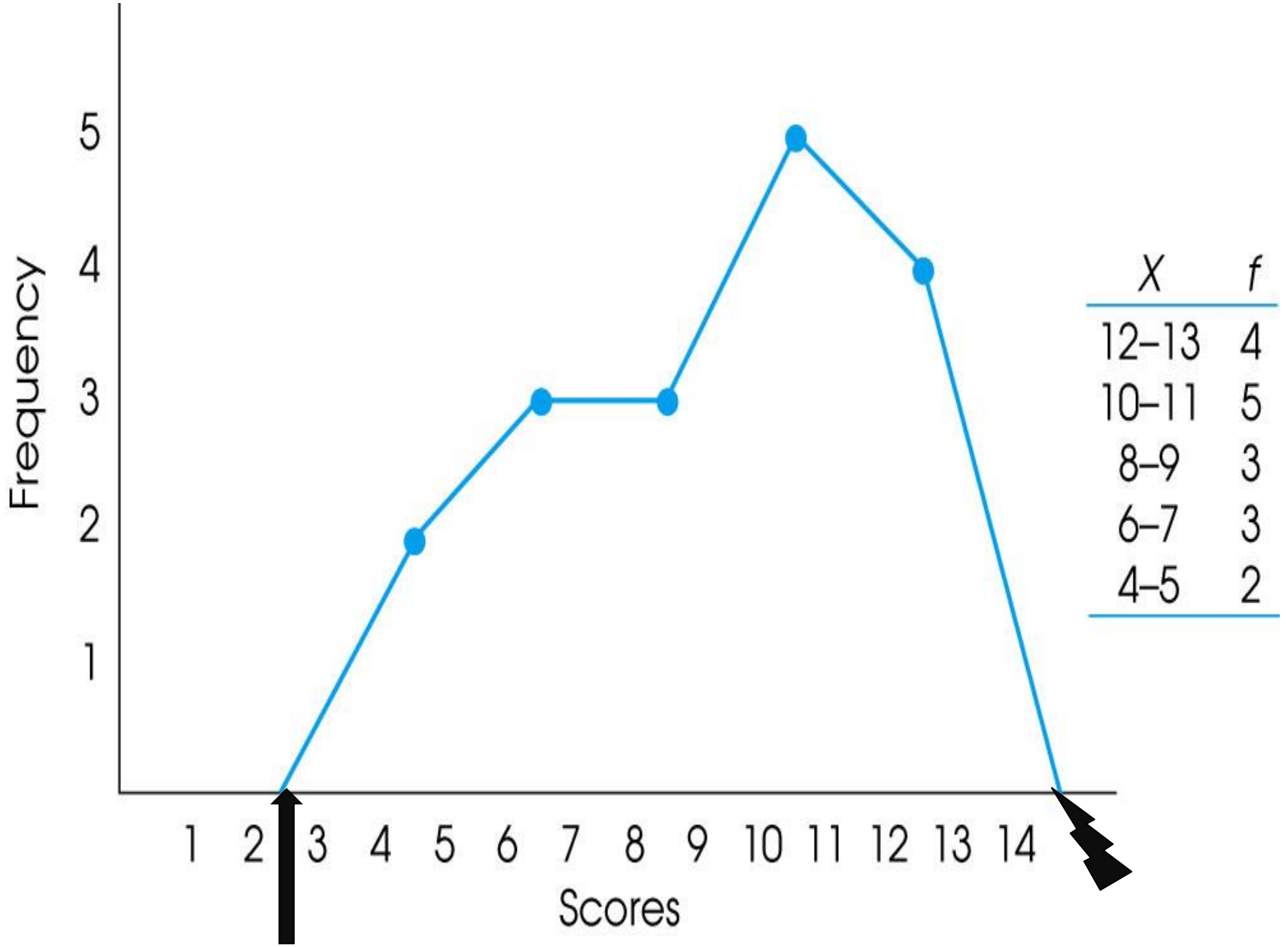


It can be constructed from histogram by taking the midpoint dot of each rectangle (class interval).

Total area of histogram graph is equal to total area under polygon graph.

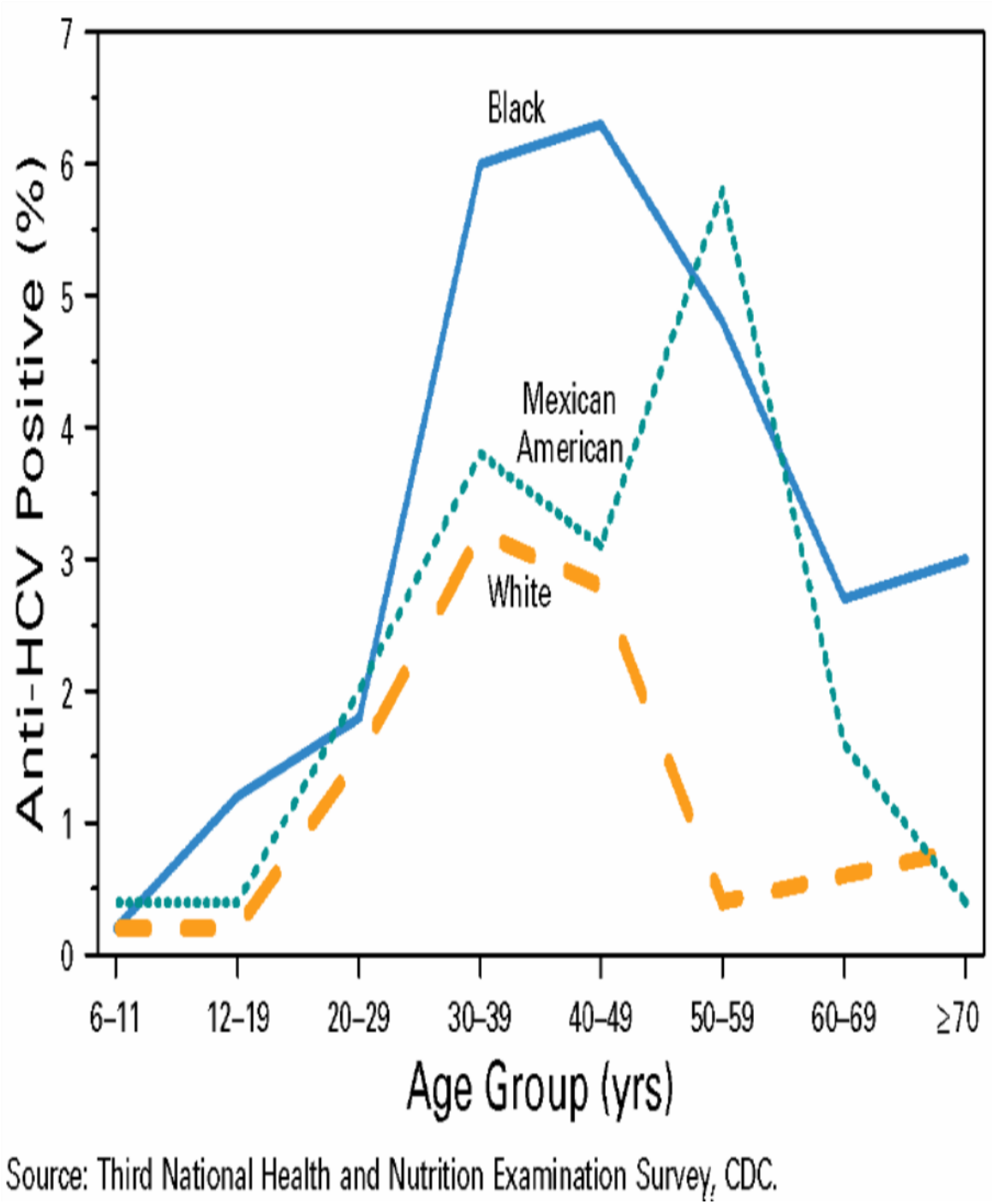
Explanation for this is by triangular method that the lines connecting midpoints of rectangles which divide them into triangles above & below the polygon; if these are gathered will give the same total area of the histogram.

Note that the polygon is brought down to the horizontal axis at the ends at points that would be the midpoints if there were an additional cell at each end of the corresponding histogram.

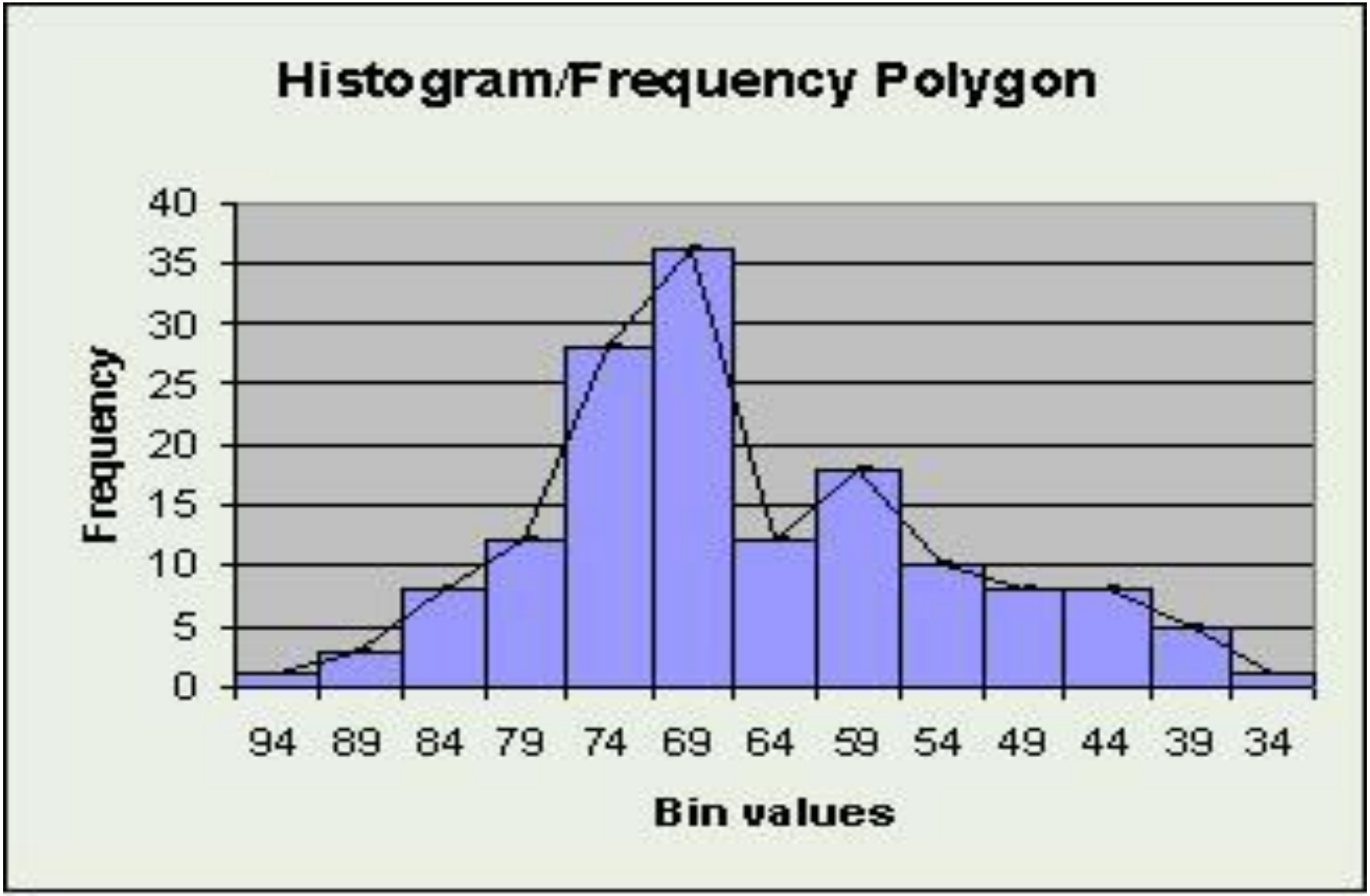


X	f
12–13	4
10–11	5
8–9	3
6–7	3
4–5	2

Prevalence of hepatitis C virus (HCV) infection by age and race/ethnicity — United States, 1988–1994

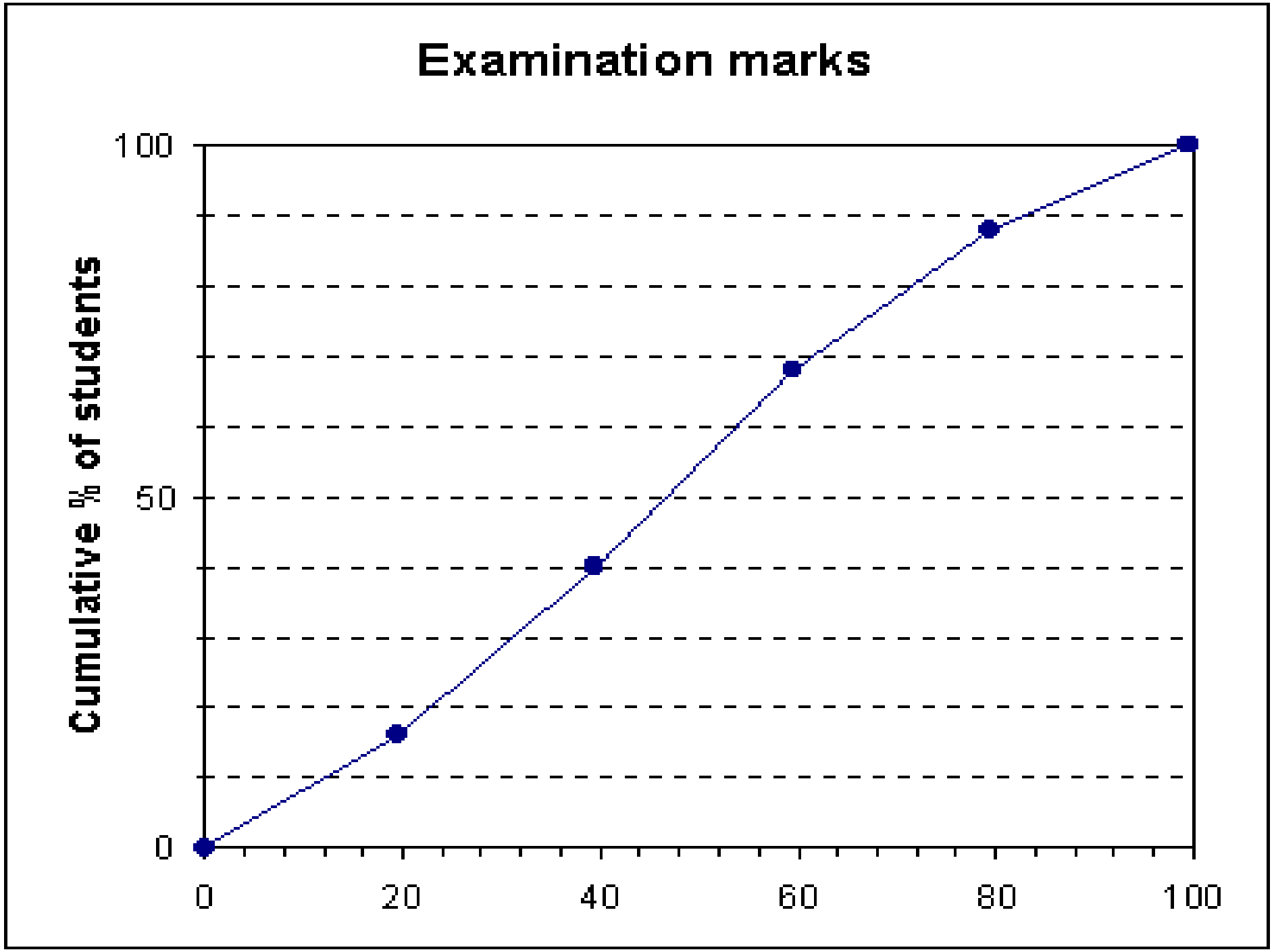


Source: Third National Health and Nutrition Examination Survey, CDC.


Histogram/Frequency Polygon
Frequency
40
35
30
25
20
15
10
5
0
94
89
84
79
74
69
64
59
54
49
44
39
34
Bin values

Cumulative frequency polygon or O" give curve

When in frequency polygon, instead of actual frequency in each class interval if cumulative frequencies are used, then it becomes cumulative frequency curve. With the help of this curve one may find out the number of observations falling below or above a specific value . This is useful for calculations of quartiles , percentiles , median , and for comparison between two or more groups.

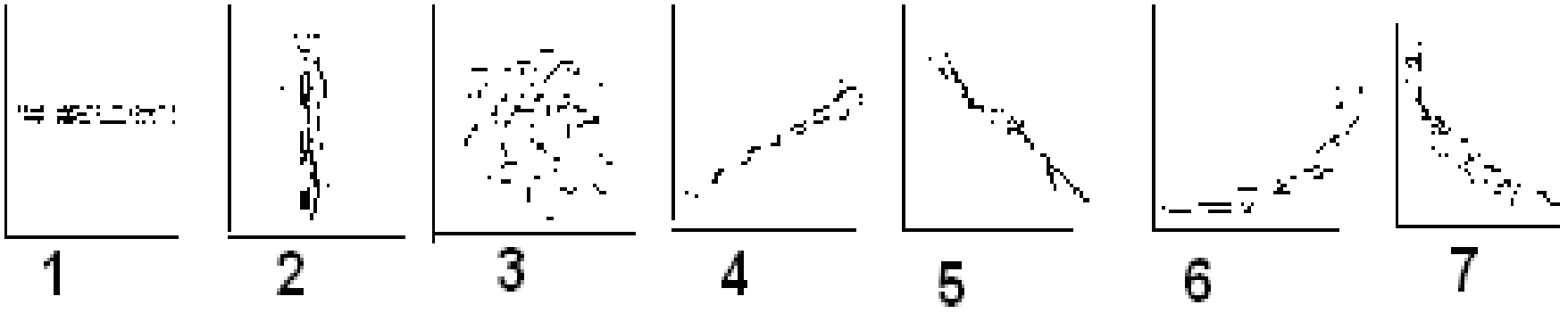

Examination marks
Cumulative % of students
100
50
0
0
20
40
60
80
100

Scatter diagram

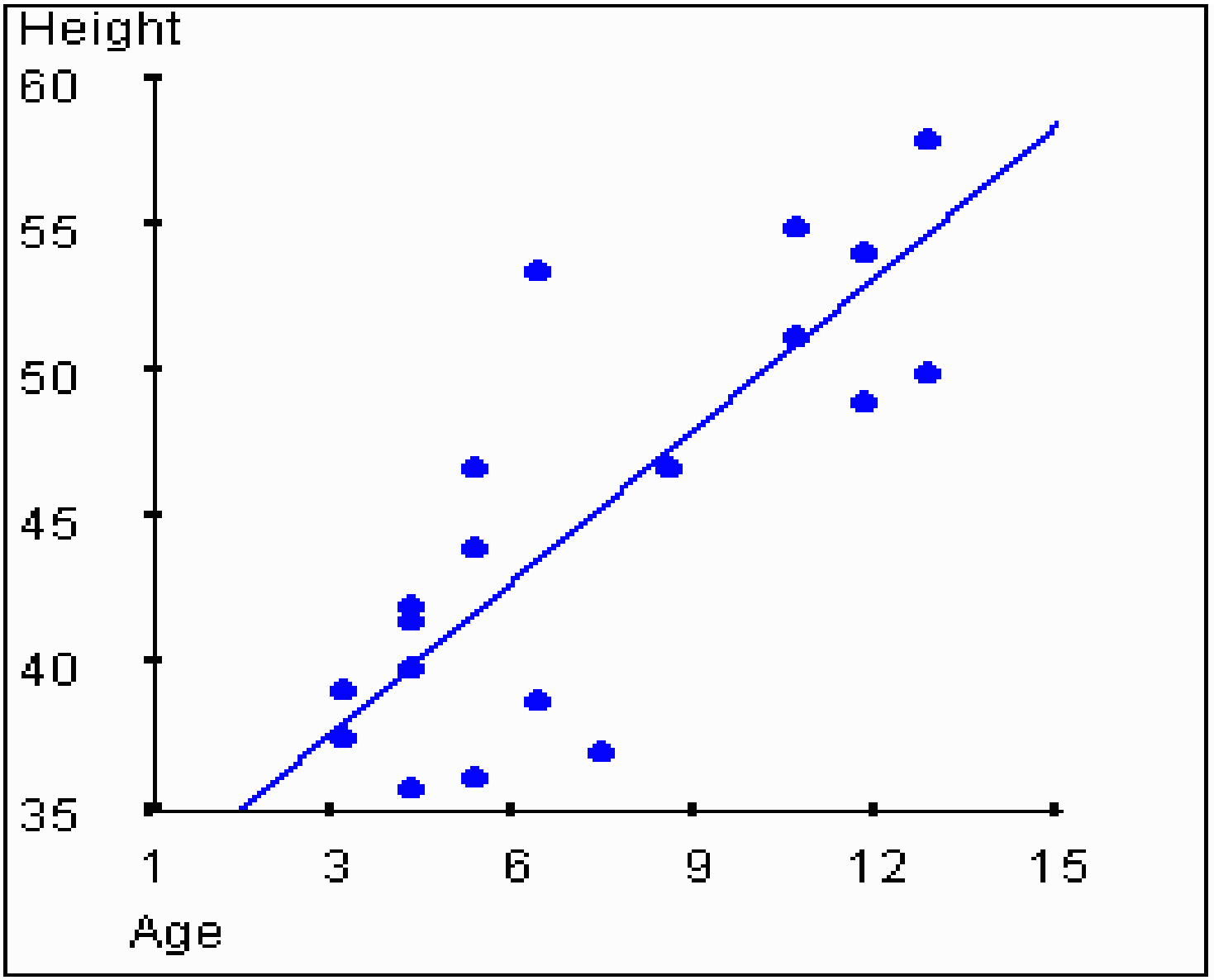
This is useful to assess the relationship between two quantitative variables .

In plotting data of this type one variable is placed on the x-axis (independent variable) and the second on the Y- axis (dependent variable).

The pattern made by these dots is indicative of a possible relationship between two variables.



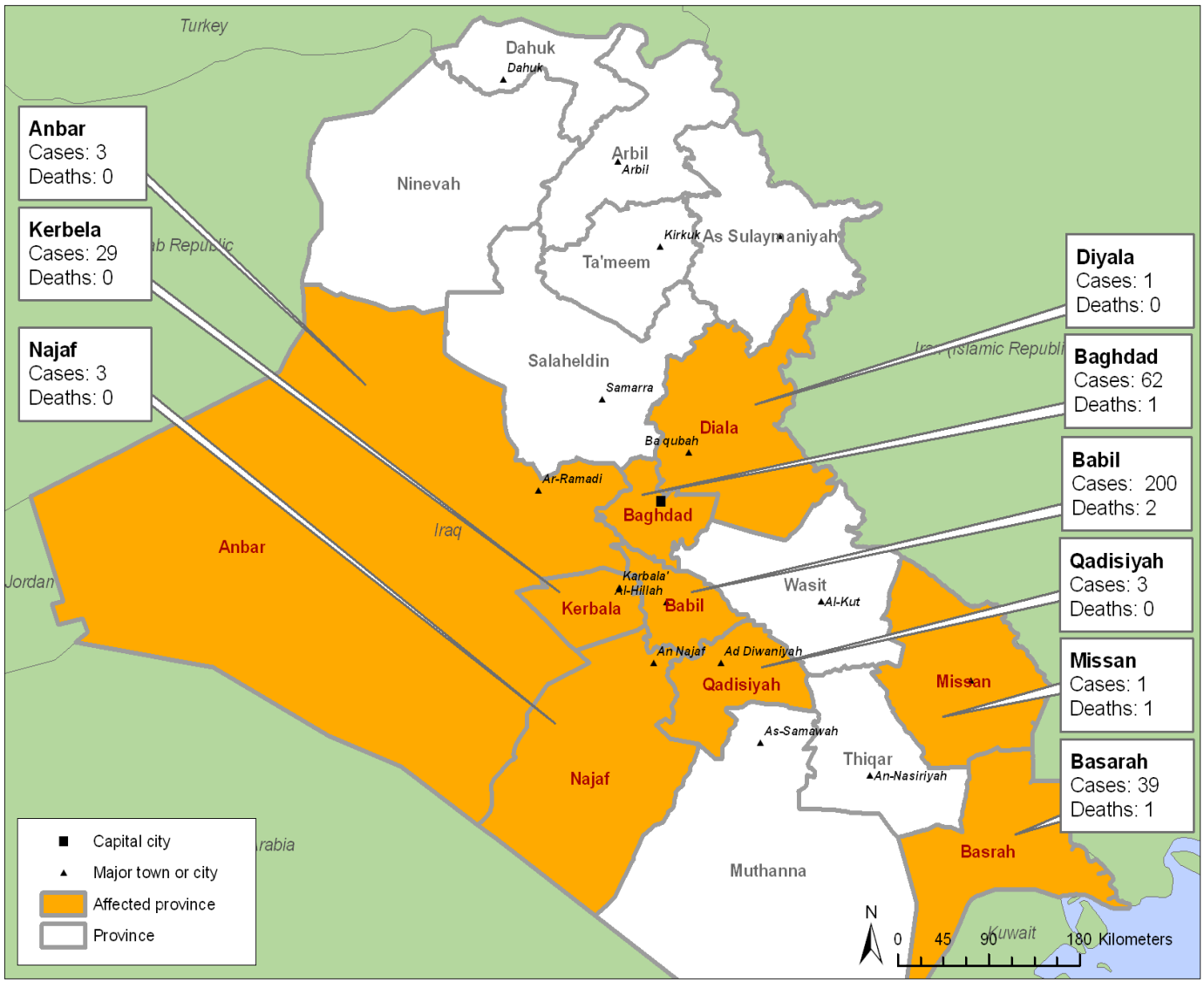
1,2&3 no relation, 4&5 +ve&–ve linear, 6&7 +ve&-ve curvi-linear


Height
60
55
50
45
40
35
1
3
6
9
12
15
Age

Map chart: geographical distribution illustrated by symbols over a map.

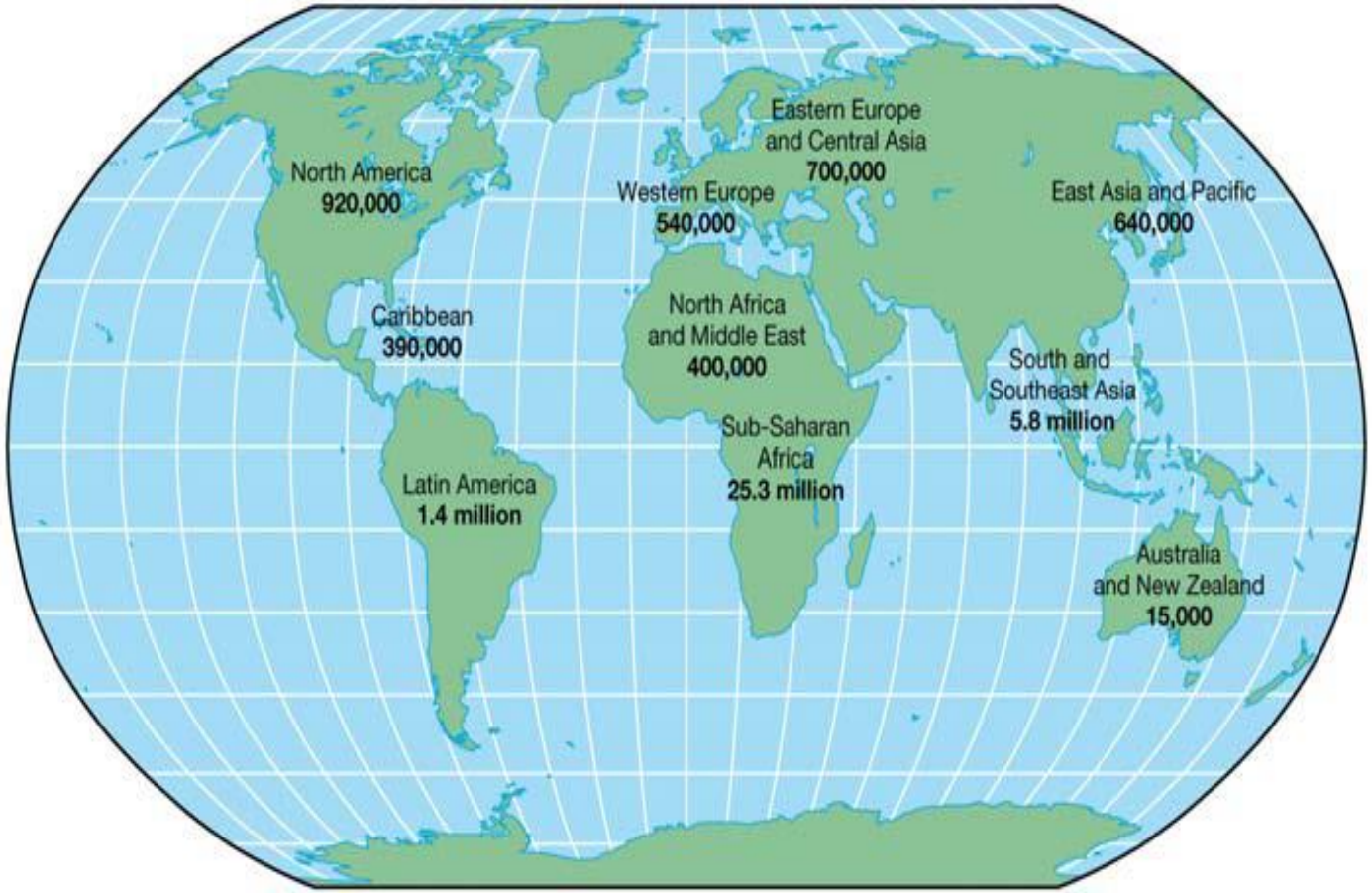
Flow chart: it shows the sequence of occurrence of a series of events.

Cholera in Iraq: 28 September 2008

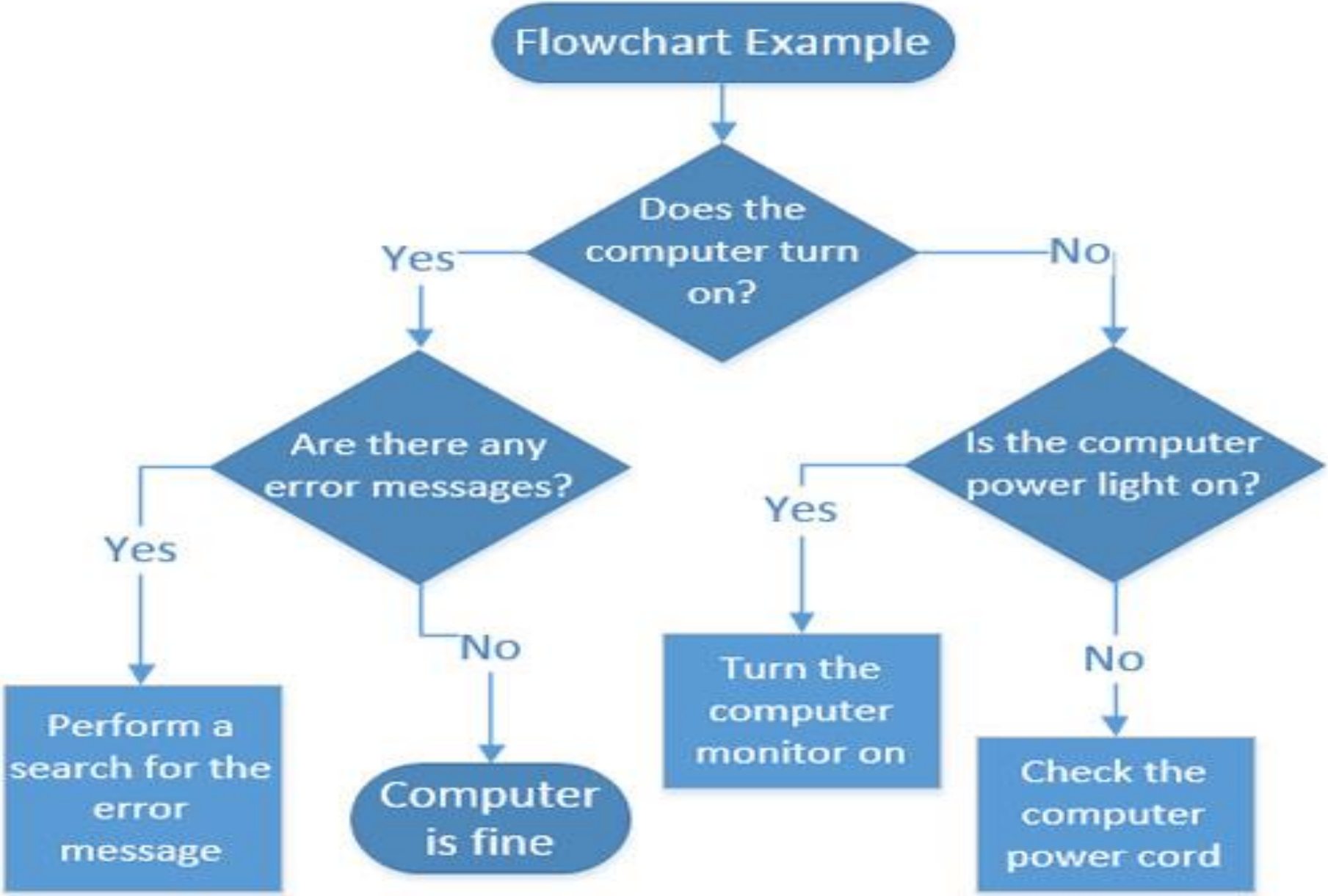


World Health Organization

Data Source: WHO Office, Iraq
Map Production: Public Health Information and Geographic Information Systems (GIS)
World Health Organization.



The AIDS epidemic - 2004


Flowchart Example
Does the computer turn on?
Yes
No
Are there any error messages?
Is the computer power light on?
Yes
No
Yes
No
Perform a search for the error message
Computer is fine
Turn the computer monitor on
Check the computer power cord
http://www.computerhope.com

Four Broad Principals of Data Presentation

Integration: Tables and graphics should be part of a " unified information flow".
Text should refer to and direct readers towards these exhibits.

Speed and Efficiency of Communication: Demonstrations should be clearly and simply presented, well-titled, and linked to the purpose of the report/memo;
The goal is efficiency of communication.

Engagement in Depth: The longer the viewer spends with an display, the more they should get out of it.
"The goal is to create a richly informative display that is dense with information, but open and accessible to the eye."

Trust value: Shows present accurate information.
They must be supported with appropriate sourcing and with all information presented correctly and understandably.

A research study has been conducted examining the number of children in the families living in a community. The following data has been collected based on a random sample of n = 30 families from the community.

2, 2, 5, 3, 0, 1, 3, 2, 3, 4, 1, 3, 4, 5, 7, 3, 2, 4, 1, 0, 5, 8, 6, 5, 4 , 2, 4, 4, 7, 6

Organize this data in a Frequency Table!

The following data has been collected based on a random sample of n = 30 patients who went to the emergency room of the clinic for Heart related problems.

The measurements are: 42, 38, 51, 53, 40, 68, 62, 36, 32, 45, 51, 67, 53, 59, 47, 63, 52, 64, 61, 43, 56, 58, 66, 54, 56, 52, 40, 55, 72, 69.