2.2 | Summarizing Quantitative data :

* We summarize quantitative data by tabular and graphical methods.

Frequency distribution : is a tabular summary of data showing the number (frequency) of item in each several non overlapping classes (categories).

* This holds for quantitative and qualilitative data
* But for quantitative data : there are three steps to define the classes for a frequency distribution :

1. Number of non overlapping classes : usually (5-20) classes
   • For small data we use 5 classes.
   • For large data we need more classes.

2. Width of each class = $\dfrac{\text{Largest data value} - \text{Smallest data value}}{\text{Number of classes}}$

   ↓ can be approximately accordingly to the preference of the person developing the frequency distribution. $9.2 \approx 10$

3. class limits : must be chosen so that each data item belongs to one and only one class.

* The lower class limit : identifies the smallest possible data value assigned to the class

* The upper class limit : = = Largest =

Class midpoint : is the value halfway between the lower and upper class limits.

Example: Consider the following data. $\qquad$ (13)

$$\begin{array}{cccc}
12 & 14 & 19 & 18 \\
15 & 15 & 18 & 17 \\
20 & 27 & 22 & 23 \\
22 & 21 & 33 & 28 \\
14 & 18 & 16 & 13
\end{array}$$

* Consider the number of class = 5
* Largest value = 33
* Smallest value = 12
* Width of each class = $\dfrac{33-12}{5} = 4.2 \approx 5$

The class (category)	Frequency	Relative frequency	Percent frequency
Lower class limit → 10 − 14 ← upper class limit	4	$\dfrac{4}{20} = \dfrac{20}{100} = 0.20$	$0.20 \times 100 = 20\%$
15 − 19	8	$\dfrac{8}{20} = \dfrac{40}{100} = 0.40$	$0.40 \times 100 = 40\%$
20 − 24	5	$\dfrac{5}{20} = \dfrac{25}{100} = 0.25$	$0.25 \times 100 = 25\%$
25 − 29	2	$\dfrac{2}{20} = \dfrac{10}{100} = 0.10$	$0.10 \times 100 = 10\%$
30 − 34	1	$\dfrac{1}{20} = \dfrac{5}{100} = 0.05$	$0.05 \times 100 = 5\%$
$n = 20$ total		Total = 1.00	Total = 100%

↖ this table is relative frequency and percent frequency distributions for the data above.

The midpoints of the classes
$$\begin{array}{ll}
10 - 14 & \rightarrow 12 \\
15 - 19 & \rightarrow 17 \\
20 - 24 & \rightarrow 22 \\
25 - 29 & \rightarrow 22 \\
30 - 34 & \rightarrow 27 \\
& \rightarrow 32
\end{array}$$

Relative frequency of class = $\dfrac{\text{Frequency of the class}}{n}$

* Percent frequency of a class = relative frequency of class × 100

* True limits $\begin{cases} \text{Upper true limits} & 14.5, 19.5, 24.5, 29.5, 34.5 \\ \text{lower true limits} & 9.5, 14.5, 19.5, 24.5, 29.5 \end{cases}$

Dot Plot : one of the simplest graphical summaries of data. The horizontal axis shows the range for the data.

10   15   20   25   30   35
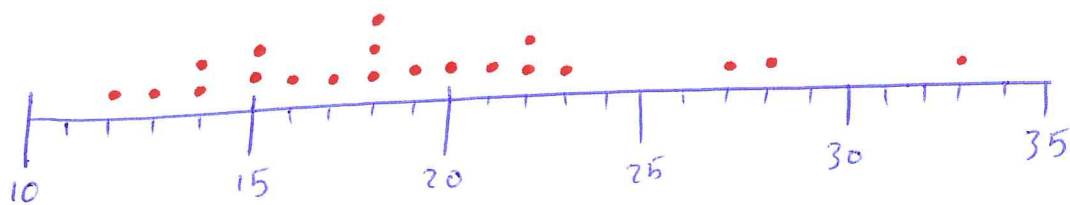
Histogram • to represent quantitative data.

• can be prepared for data summarized in frequency, relative frequency or percent frequency distribution.

→ The variable of interest on the horizontal axis and the frequency (or relative or percent frequency) on vertical axis.
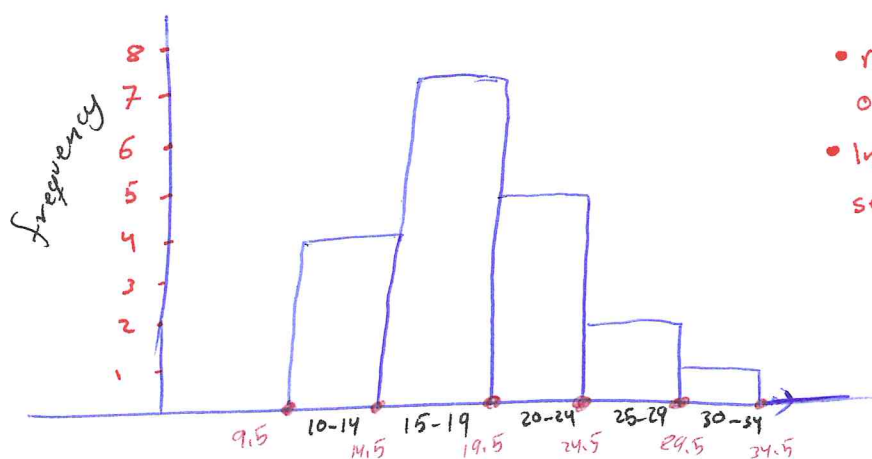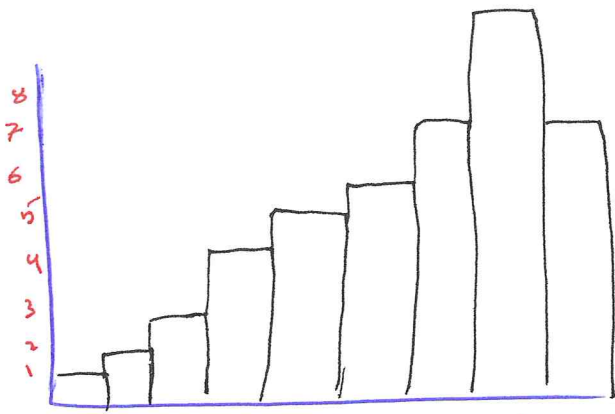

• rectangles are on touch one another in histogram
• In Bar, there is a separation.

9.5 10-14 15-19 20-24 25-29 30-34
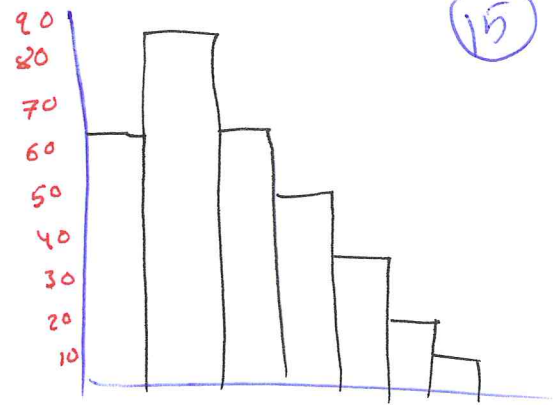   14.5   19.5   24.5   29.5   34.5

→ Histogram provides information about the shape or form of a distribution.

* A histogram is said to be skewed to the left if its tail extends farther to the left
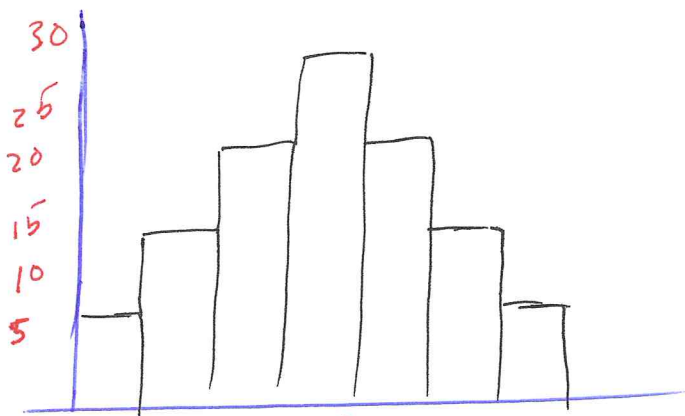
* "   "   "   =   =   =   =   =   =   right   =   =   =   =   =   = right.
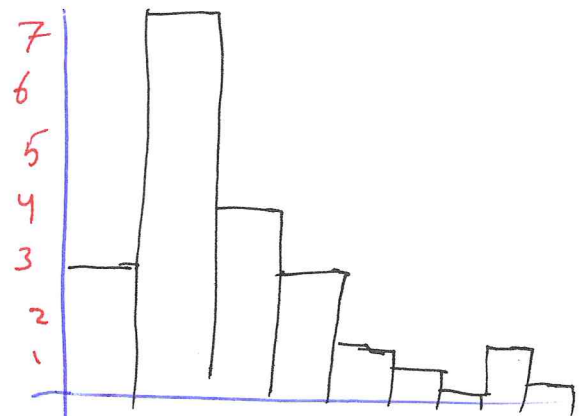
Moderately skewed left
Jlus1

Moderately skewed
Right

Symmetric

Highly skewed Right

Cumulative distributions:

the cumulative frequency distribution shows the number of data with values less than or equal to the upper class limit of each class.

Cumulative classes	Cumulative frequency	Cumulative relative frequency	Cumulative percent freq.
less than or equal 14	4	$4/20 = \frac{20}{100} = 0.20$	$0.2 \times 100 = 20\%$
less than or equal 19	12	$\frac{12}{20} = \frac{60}{100} = 0.60$	$0.60 \times 100 = 60\%$
less than or equal 24	17	$\frac{17}{20} = \frac{85}{100} = 0.85$	$0.85 \times 100 = 85\%$
less than or equal 29	19	$\frac{19}{20} = \frac{0.95}{100} = 0.95$	$0.95 \times 100 = 95\%$
less than or equal 34	20 = n	$\frac{20}{20} = \frac{100}{100} = 1.00$	$1.00 \times 100 = 100\%$

* Note that the cumulative frequency distribution shows the (16) proportion of data item.

* Note that the cumulative percent frequency distribution shows the percentage of data item with values less than or equal to the upper limit of each class.

Ogive : is the graph of a cumulative distribution.
 • Data values are on horizontal axis.
 • Cumulative frequency or cumulative relative frequency or cumulative percent frequencies on the vertical axis.