Chapter Two
Presentation of a Statistical Data

1. Introduction
2. Organizing Data
3. Histograms, Frequency Polygons, and Ogives
4. Other Types of Graphs

Chapter Two

Presentation of a Statistical Data

1. Introduction

- Gathering data for a particular variable under study is the primary task for presenting the data.
- The data must be organized in some meaningful way. The most convenient method of organizing data is to construct a frequency distribution.
- Then data must be presented to be understood by those who will benefit from reading the study.
- The most useful method of presenting the data is by constructing **statistical charts and grap**hs.

2. Organizing Data

- Information can be obtained from looking at raw data (Table 1), the data to be more understandable, they should be organized. One of the comment statistic methods are using so called *frequency distribution (Table 2)*.
  - A **frequency distribution** is the organization of raw data in a table form, using **classes and frequencies**.

Table 1: Raw data

49	57	38	73	81
74	59	76	65	69
54	56	69	68	78
65	85	49	69	61
48	81	68	37	43
78	82	43	64	67
52	56	81	77	79
85	40	85	59	80
60	71	57	61	69
61	83	90	87	74

Table 2: Frequency distribution table.

Class limits	Tally	Frequency
35–41	///	3
42–48	///	3
49–55	////	4
56–62	𝍸 //	10
63–69	𝍸 //	10
70–76	𝍸	5
77–83	𝍸 //	10
84–90	𝍸	5
	Total	50

Classes

Frequencies

❑ Grouped Frequency Distributions or Frequency Distributions Table

When the range of the data is large or huge, the data must be grouped into classes with the frequency of each class as shown in Table 2.

➢ Procedure for Constructing the Frequency Distribution Table

There are some concepts need to be explained as shown in the following distribution frequency table (Table 3).

- The values of the first class are called *class limits such as (24-30).* The **lower class limit (**24) represents the smallest data value that can be included in the class. The **upper class limit** (30) represents the largest data value that can be included in the class.
- The numbers in the second column are called **class boundaries.** These numbers are used to separate the classes so that there are no gaps in the frequency distribution.

Table 3: Frequency distribution table.

Class limits	Class boundaries	Tally	Frequency
24–30	23.5–30.5	///	3
31–37	30.5–37.5	/	1
38–44	37.5–44.5	𝍸	5
45–51	44.5–51.5	𝍸 ////	9
52–58	51.5–58.5	𝍸 /	6
59–65	58.5–65.5	/	1
			25

Note: The *class limits should have the same decimal place value as the data, but the class boundaries should have one additional place value and end in a 5.*

For example: the boundaries limits for the classes (31–37) & (7.8–8.8), are:

Lower limit -0.5 = 31 - 0.5 = 30.5 lower boundary

Upper limit + 0.5 = 37 + 0.5 = 37.5 upper boundary

Lower limit -0.05 = 7.8 - 0.05 = 7.75 lower boundary

Upper limit +0.05 = 8.8 + 0.05 = 8.85 upper boundary

Class limits	Class boundaries	Tally	Frequency
24–30	23.5–30.5	///	3
31–37	30.5–37.5	/	1
38–44	37.5–44.5	𝍸	5
45–51	44.5–51.5	𝍸 ////	9
52–58	51.5–58.5	𝍸 /	6
59–65	58.5–65.5	/	1
			25

- ***Class width (C_w)*** *is the range between upper and lower limit of the same class.*

C_w = the lower (or upper) class limit of one class - the lower (or upper) class limit of the next class.

For example: *the class width of Table 3 is:*

31-24 = 7 OR 37-30 = 7 OR 23.5-30.5 = 7 OR 37.5-30.5 = 7

- ***Number of classes are between 5 and 20 classes.***
- **The class midpoint X_m is**

$$X_m = \frac{lower\ boundary + upper\ boundary}{2}$$

OR $$X_m = \frac{lower\ limit + upper\ limit}{2}$$

Example:

$$\frac{24 + 30}{2} = 27 \quad \text{or} \quad \frac{23.5 + 30.5}{2} = 27$$

Example 1: These data represent the record high temperatures in degrees Fahrenheit (F) for each of the 50 states. Construct a grouped Frequency distribution for the data using 7 classes.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Solution:

1. Find the highest value and lowest value: H = 134 and L = 100.
2. Find the range: R = highest value - lowest value = $H - L$; R = 134 – 100 = 34
3. Select the number of classes (5-20); n = 7.
4. Find the class width; $C_w = \frac{R}{n} = \frac{34}{7}$ **= 4.9 ≈ 5** **OR** **4.0**
5. Select a starting point for the lowest class limit *= lowest value or less (100 or 99).*
6. *Determine the lower limits of the other class = Lower limit +* C_w *= 100 + 5 = 105, 110, 115, etc.*
7. *Determine the Upper limits of the first class =*
   lower limit (2nd class) – 1 (one unit) = ***105-1=104***
8. *Determine the upper limits of the other class = lower limit +* C_w *= 104 + 5 = 109, 114, 119, etc.*

9. Find the class boundaries: by subtracting 0.5 from each lower class limit and adding 0.5 to each upper class limit: First class : 99.5–104.5, second class: 104.5–109.5, etc.
10. Tally the data.
11. Find the numerical frequencies from the tallies.

Table 4: Frequency distribution table.

Class limits	Class boundaries	Tally	Frequency
100–104	99.5–104.5	//	2
105–109	104.5–109.5	𝍸 ///	8
110–114	109.5–114.5	𝍸 𝍸 𝍸 ///	18
115–119	114.5–119.5	𝍸 𝍸 ///	13
120–124	119.5–124.5	𝍸 //	7
125–129	124.5–129.5	/	1
130–134	129.5–134.5	/	1
			$n = \Sigma f = 50$

?

=

12. Further Calculations:

- The cumulative frequency distribution: It is a distribution that shows the number of data values less or higher than or equal to a specific value (usually an upper or lower boundary).

Class limits	Class boundaries	Tally	Frequency
100–104	99.5–104.5	//	2
105–109	104.5–109.5	𝍸 ///	8
110–114	109.5–114.5	𝍸 𝍸 𝍸 ///	18
115–119	114.5–119.5	𝍸 𝍸 ///	13
120–124	119.5–124.5	𝍸 //	7
125–129	124.5–129.5	/	1
130–134	129.5–134.5	/	1
			$n = \Sigma f = 50$

The cumulative frequency

Ascending cumulative frequency (Less than X)

Ex: Less than 99.5 = 0

Less than 104.5 = 0 + 2 = 2

Less than 19.5 = 0+2+8+18+13 = 31

	Cumulative frequency
Less than 99.5	0
Less than 104.5	2
Less than 109.5	10
Less than 114.5	28
Less than 119.5	41
Less than 124.5	48
Less than 129.5	49
Less than 134.5	50

Table 5: Ascending cumulative frequency distribution table.

Descending cumulative frequency (Greater than X)

Ex: Greater than 99.5 = 50

Greater than 104.5 = 50-2 = 48

Greater than 114.5 = 50 – 18 -8 -2 = 22

Note: Cumulative frequencies are used to show how many data values are accumulated up or down to and including a specific class.

	Cumulative frequency
Greater than 99.5	50
Greater than 104.5	48
Greater than 109.5	40
Greater than 114.5	22
Greater than 119.5	9
Greater than 124.5	2
Greater than 129.5	1
Greater than 134.5	0

Table 6 Descending cumulative frequency distribution table.

➢ Briefly

The following guides line steps can be used for constructing the frequency distribution table:

Procedure Table

Constructing a Grouped Frequency Distribution

Step 1 Determine the classes.

- Find the highest and lowest values.
- Find the range.
- Select the number of classes desired.
- Find the width by dividing the range by the number of classes and rounding up.
- Select a starting point (usually the lowest value or any convenient number less than the lowest value); add the width to get the lower limits.
- Find the upper class limits.
- Find the boundaries.

Step 2 Tally the data.

Step 3 Find the numerical frequencies from the tallies, and find the cumulative frequencies.

3. Histograms, Frequency Polygons, and Ogives

- Statistical graphs can be used to describe the data set or to analyze it.
- The purposes of using graphs are:

✓ to discuss an issue,
✓ reinforce a critical point
✓ summarize a data
✓ discover the trend or pattern in a situation over a period of time.

The three most commonly used graphs are:

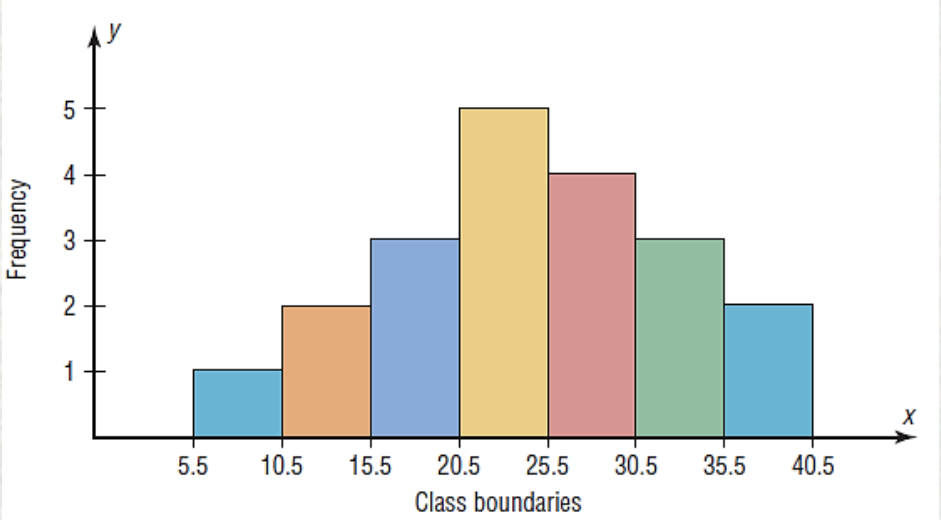
1. The histogram.
2. The frequency polygon.
3. The cumulative frequency graph, or ogive.

1. Histogram

The **histogram** is a graph that displays the data by using contiguous vertical bars (unless the frequency of a class is 0) of various heights to represent the frequencies of the classes.

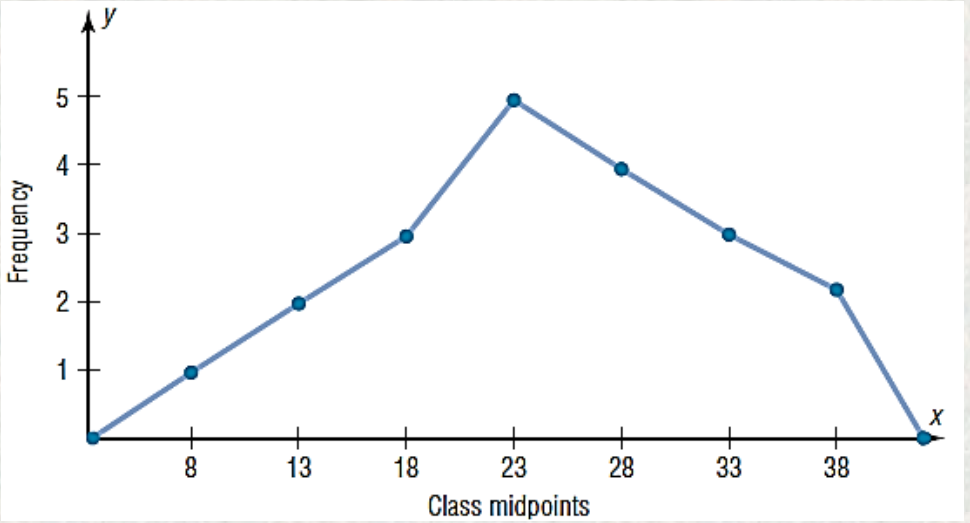
Example:

Class boundaries	Frequency
99.5–104.5	2
104.5–109.5	8
109.5–114.5	18
114.5–119.5	13
119.5–124.5	7
124.5–129.5	1
129.5–134.5	1

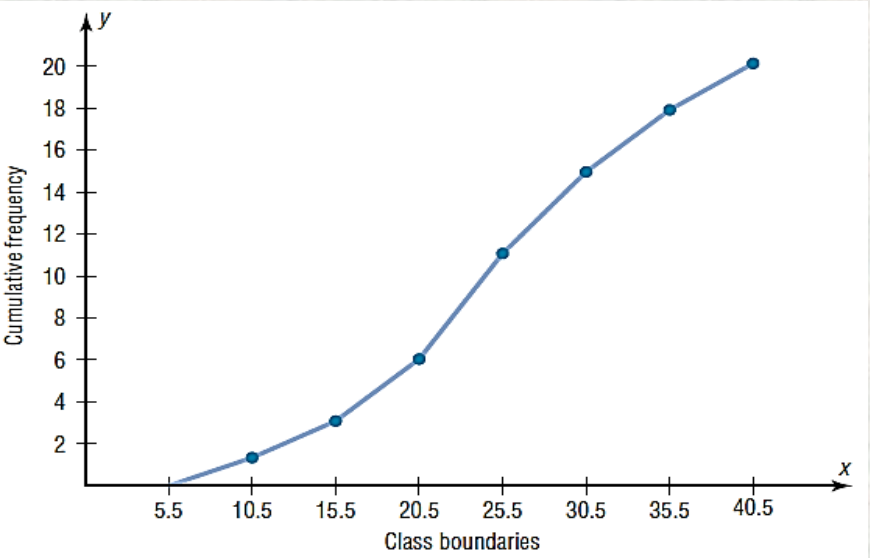


2. Frequency Polygon

The frequency **polygon** is a graph that displays the data by using lines that connect points plotted for the frequencies at the midpoints of the classes. The frequencies are represented by the heights of the points.



3. Ogive



This type of graph is called the **cumulative frequency graph**, or **Ogive**. The cumulative frequency is the sum of the frequencies accumulated up to the upper boundary of a class in the distribution.

Example:

Construct a histogram, polygon and Ogive to represent the data shown for the record high temperatures.

Class boundaries	Frequency
99.5-104.5	2
104.5-109.5	8
109.5-114.5	18
114.5-119.5	13
119.5-124.5	7
124.5-129.5	1
129.5-134.5	1

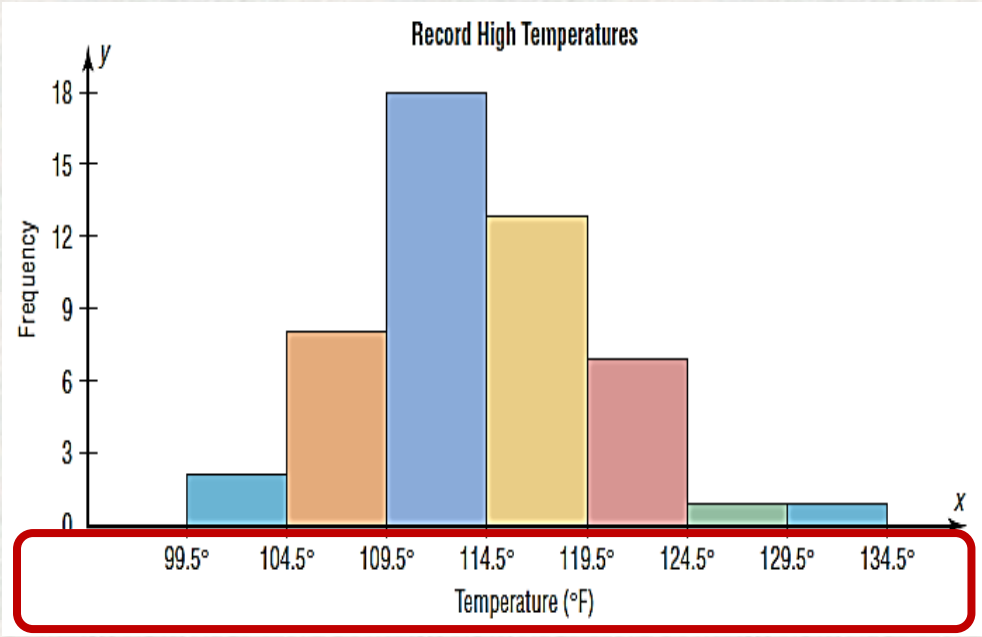
Solution

1. Find the midpoints of each class. Recall that midpoints are found by adding the upper and lower boundaries and dividing by 2.

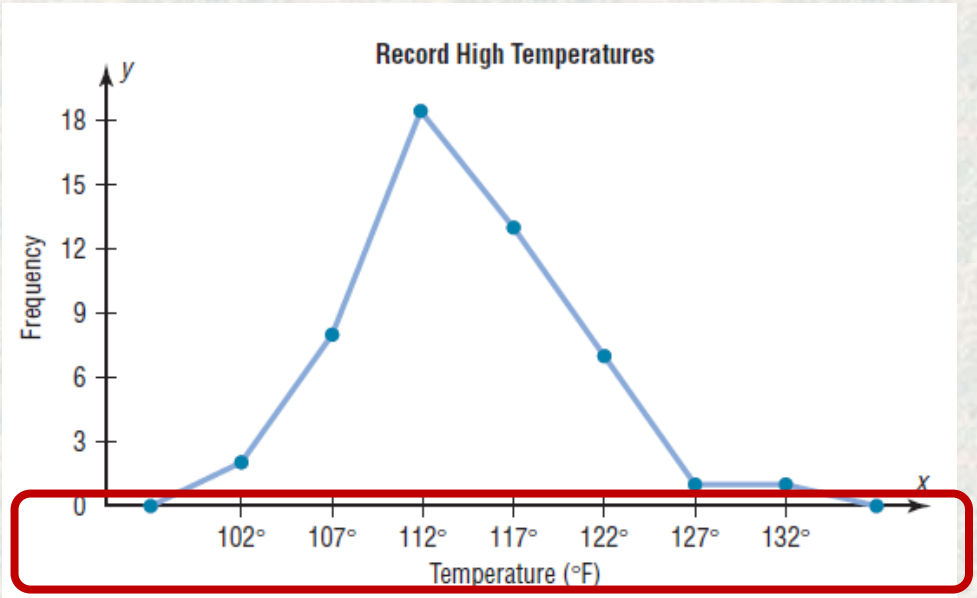
$$\frac{99.5 + 104.5}{2} = 102 \qquad \frac{104.5 + 109.5}{2} = 107$$

2. Draw and label the x and y axes. The x axis is always the horizontal axis, and the y axis is always the vertical axis.
3. Using the frequencies as the heights (Y-axes), and midpoints or boundary limits as (X-axis).

Class boundaries	Midpoints	Frequency
99.5-104.5	102	2
104.5-109.5	107	8
109.5-114.5	112	18
114.5-119.5	117	13
119.5-124.5	122	7
124.5-129.5	127	1
129.5-134.5	132	1



Boundaries limits



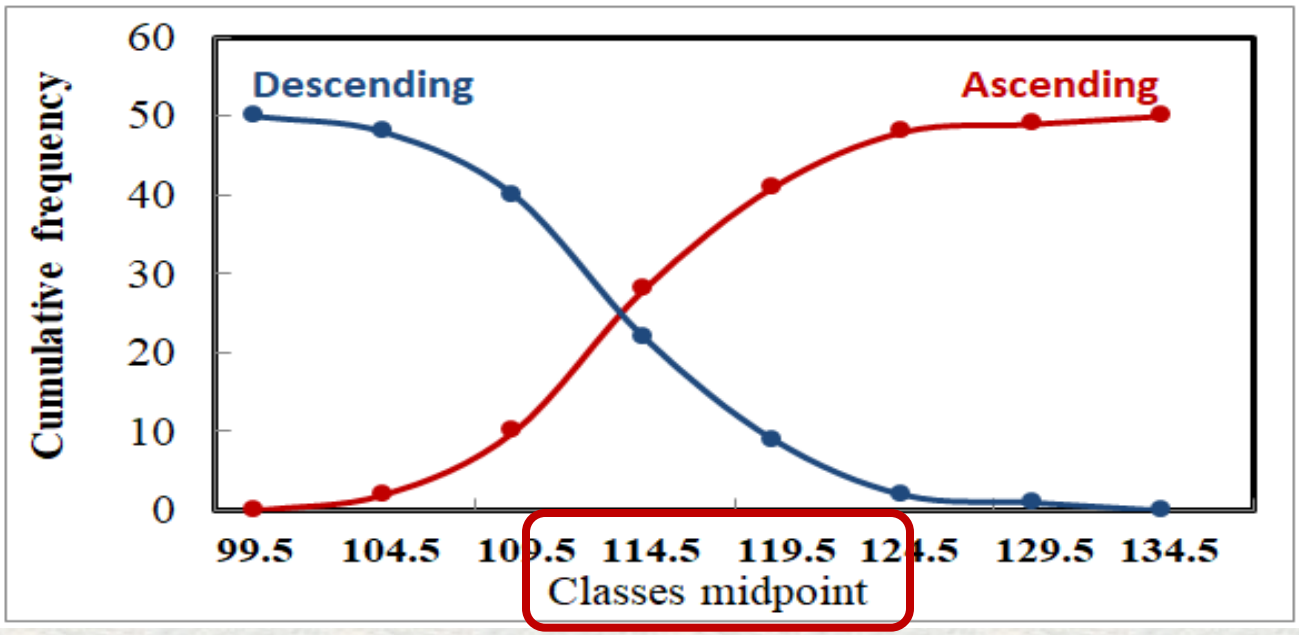
Classes midpoint

4. Find the cumulative frequency for each class.

Class boundaries	Midpoints	Frequency
99.5-104.5	102	2
104.5-109.5	107	8
109.5-114.5	112	18
114.5-119.5	117	13
119.5-124.5	122	7
124.5-129.5	127	1
129.5-134.5	132	1

Ascending cumulative frequency	
Less than 99.5	0
Less than 104.5	2
Less than 109.5	10
Less than 114.5	28
Less than 119.5	41
Less than 124.5	48
Less than 129.5	49
Less than 134.5	50

Descending cumulative frequency	
Greater than 99.5	50
Greater than 104.5	48
Greater than 109.5	40
Greater than 114.5	22
Greater than 119.5	9
Greater than 124.5	2
Greater than 129.5	1
Greater than 134.5	0



➢ **Briefly :**

The following guides line steps can be used for constructing the frequency distribution table:

Procedure Table

Constructing Statistical Graphs

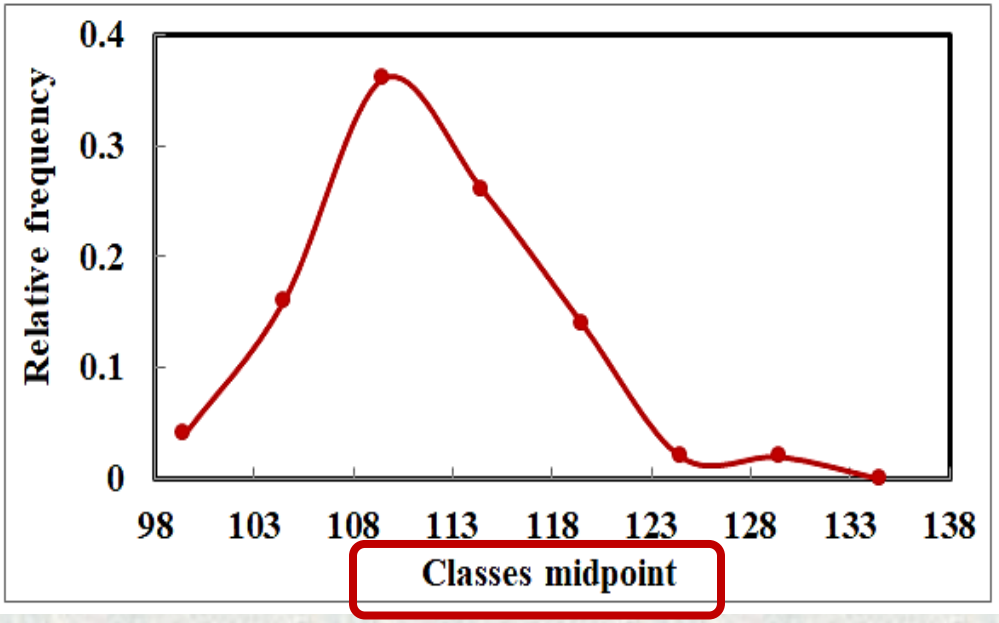
Step 1	Draw and label the x and y axes.
Step 2	Choose a suitable scale for the frequencies or cumulative frequencies, and label it on the y axis.
Step 3	Represent the class boundaries for the histogram or ogive, or the midpoint for the frequency polygon, on the x axis.
Step 4	Plot the points and then draw the bars or lines.

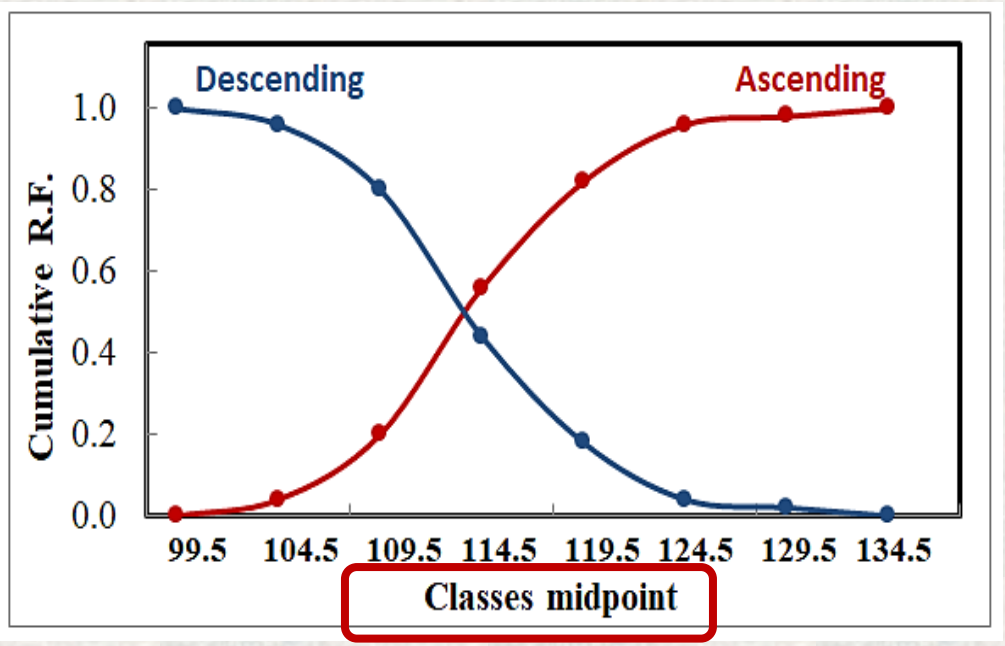
4. Relative frequency

- The histogram, the frequency polygon, and the ogive shown previously were constructed by using frequencies in terms of the raw data. These distributions can be converted to distributions using *proportions* instead of raw data as frequencies. These types of graphs are called **relative frequency graphs.**
- **Relative frequency** (F_i) can be calculated by dividing the frequency for each class (f_i) by the total of the frequencies Σf_i. The sum of the relative frequencies will always be 1.

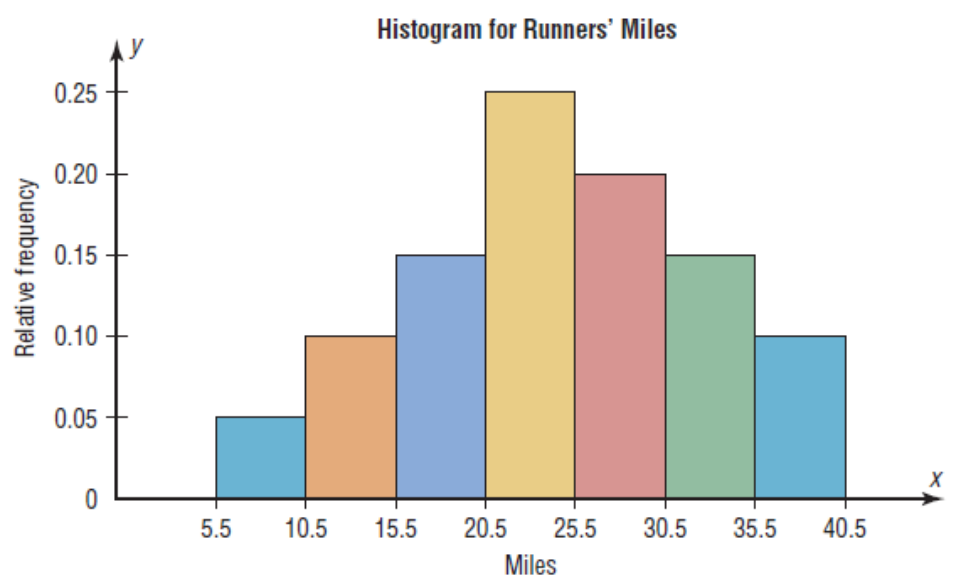
$$F_i = \frac{f_i}{\sum f_i}$$

Class boundaries	Midpoints	Frequency	Relative frequency	Ascending C.R.F.	Descending C.R.F.
99.5-104.5	102	2	0.04	0.00	1.00
104.5-109.5	107	8	0.16	0.04	0.96
109.5-114.5	112	18	0.36	0.20	0.80
114.5-119.5	117	13	0.26	0.56	0.44
119.5-124.5	122	7	0.14	0.82	0.18
124.5-129.5	127	1	0.02	0.96	0.04
129.5-134.5	132	1	0.02	0.98	0.02
				1.00	0.00

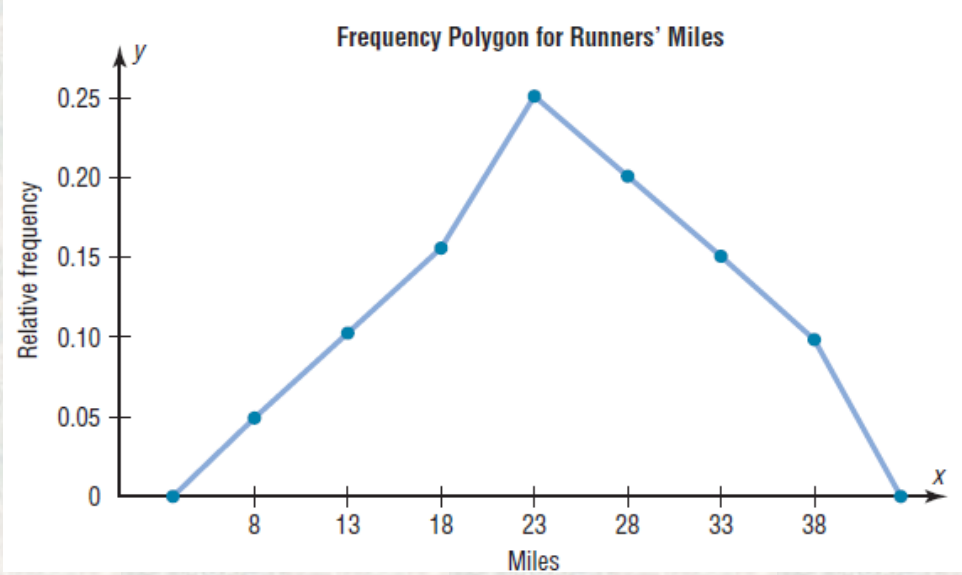




Relative frequency for histogram

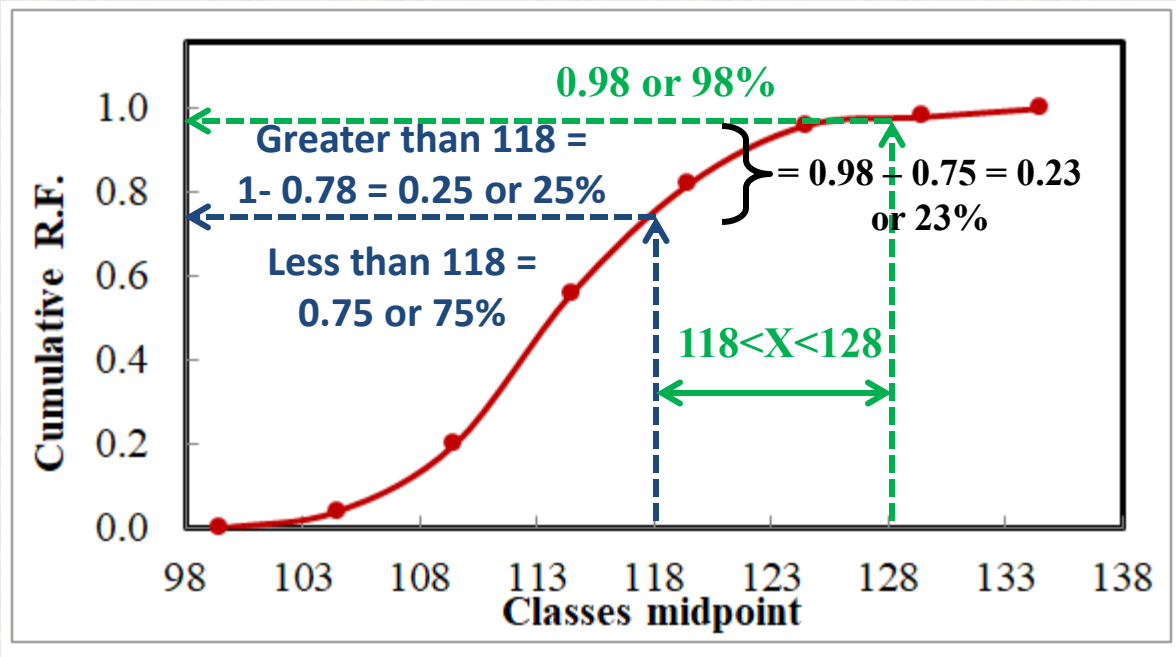


Relative Polygon for histogram



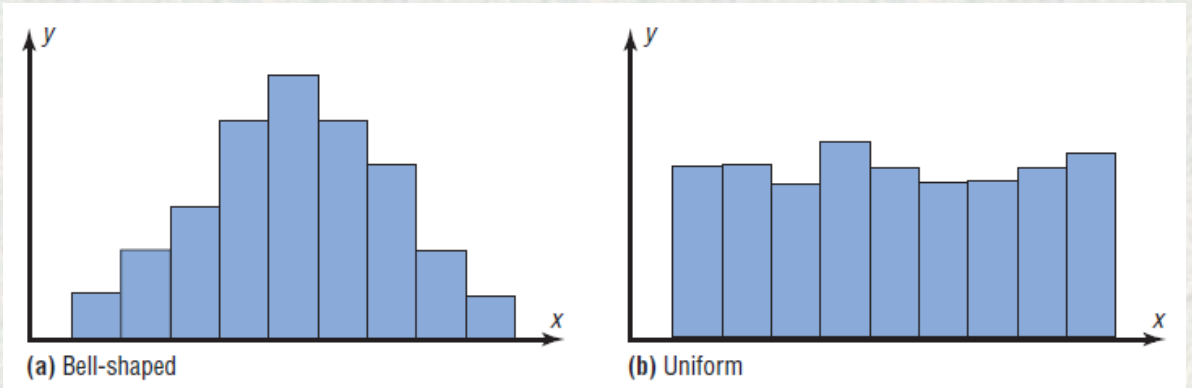
Examples for calculation

Relative frequency for histogram



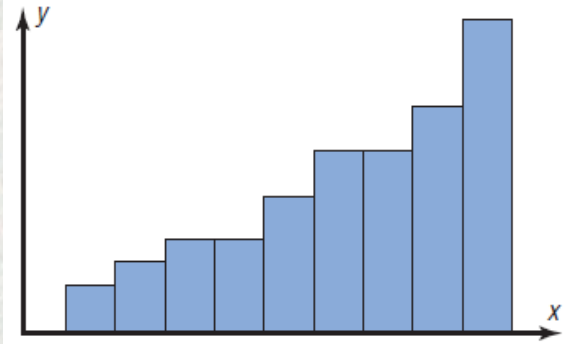
4. Distribution Shapes

The shape of a distribution determines the appropriate statistical methods used to analyze the data.

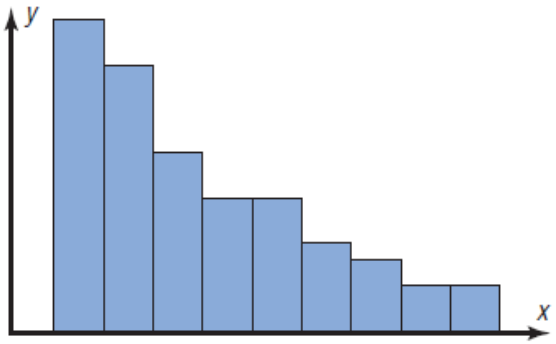


(a) Bell-shaped

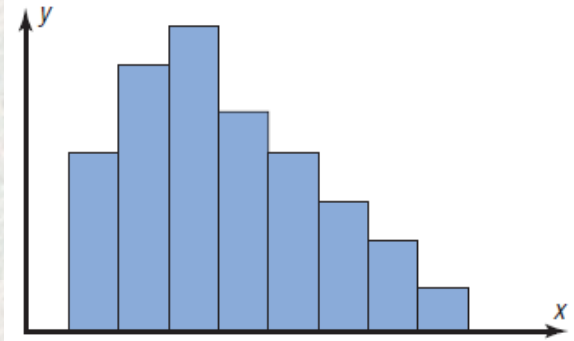
(b) Uniform



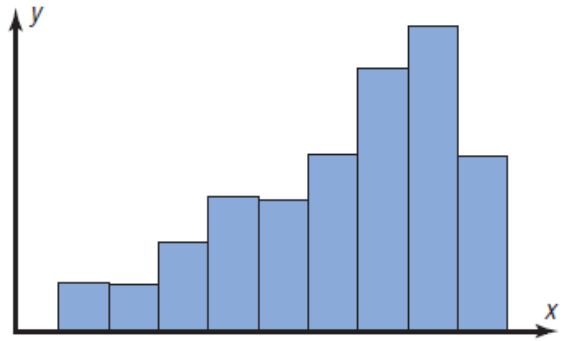
(c) J-shaped



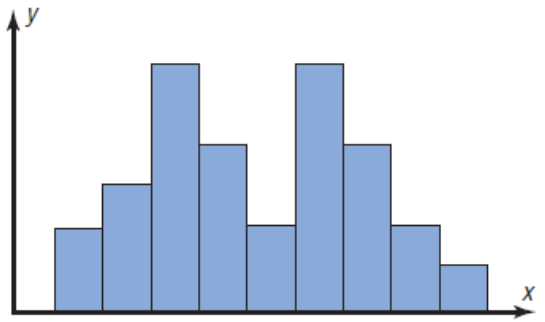
(d) Reverse J-shaped



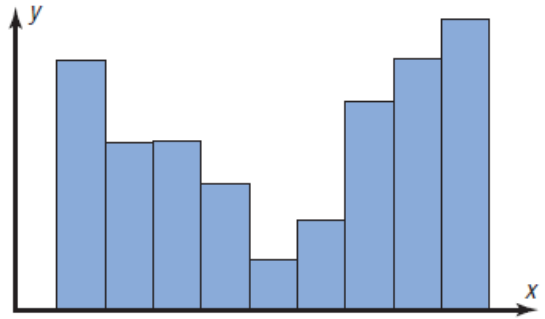
(e) Right-skewed



(f) Left-skewed

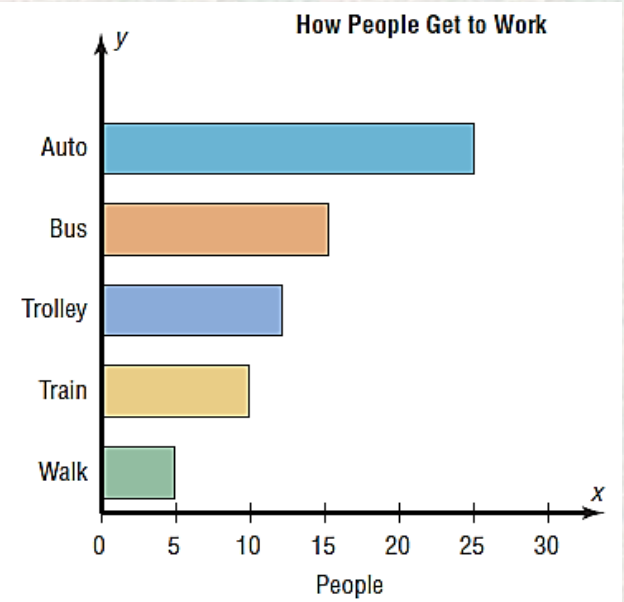


(g) Bimodal



(h) U-shaped

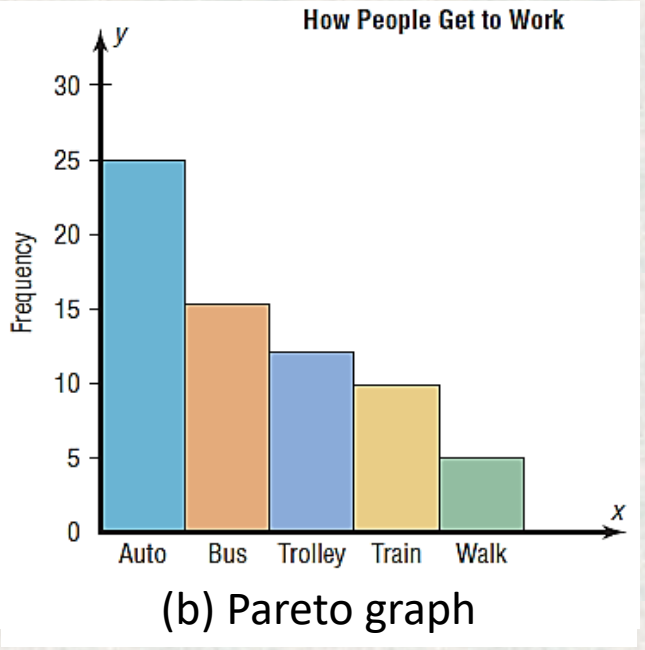
5. Other Types of Graphs



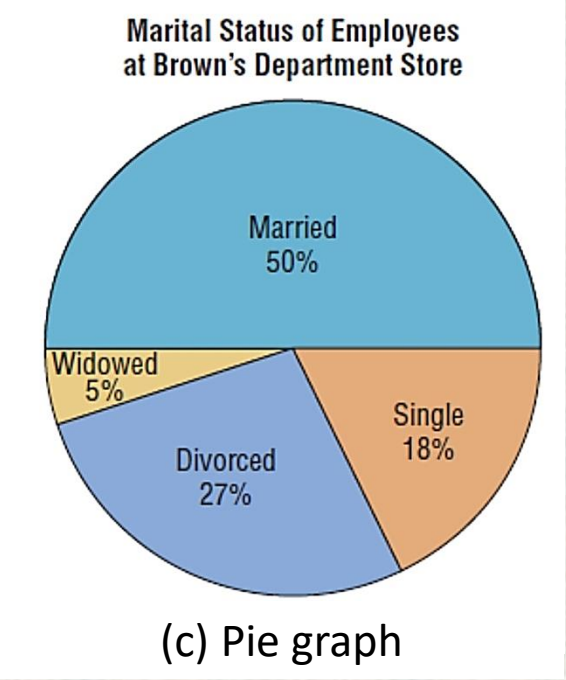
(a) Bar graph

(a) A **bar graph** represents the data by using vertical or horizontal bars whose heights or lengths represent the frequencies of the data.

(b) A **Pareto chart** is used to represent a frequency distribution for a categorical variable, and the frequencies are displayed by the heights of vertical bars, which are arranged in order from highest to lowest.



(b) Pareto graph



(c) Pie graph

(c) A **pie graph** is a circle that is divided into sections or wedges according to the percentage of frequencies in each category of the distribution.

Thank You

Any Questions?