4. Frequency Distributions and Graphs

When conducting a statistical study, the researcher must gather data for the variable under study. For example, if a researcher wishes to study the types of cement activity , he or she has to gather the data from various projects, .To describe situations, draw conclusions, or make inferences about events , the researcher must organize the data in some meaningful way. The most convenient method of organizing data is to construct a **frequency distribution.** After organizing the data, the researcher must present them so they can 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 graphs. There are many different types of charts and graphs, and each one has a specific purpose. This chapter explains how to organize data by constructing frequency distributions and how to present the data by constructing charts and graphs.

4.1. Organizing Data:

Suppose a researcher wished to do a study on the ages of the top 50 wealthiest people in the world. The researcher first would have to get the data on the ages of the people. In this case, these ages are listed in Forbes Magazine. When the data are in original form, they are called raw data and are listed next.

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

Since little information can be obtained from looking at raw data, the researcher organizes the data into what is called a frequency distribution. A frequency distribution consists of classes and their corresponding frequencies. Each raw data value is placed into a quantitative or qualitative category called a class.

The frequency of a class then is the number of data values contained in a specific class. A frequency distribution is shown for the preceding data set.

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

Example 1: Twenty-five army inductees were given a blood test to determine their blood type. The data set is

A	B	B	AB	O
O	O	B	AB	B
B	B	O	A	O
A	O	O	O	AB
AB	A	O	B	A

Construct a frequency distribution for the data.

Solution:

A Class	B Tally	C Frequency	D Percent
A	𝍸	5	20
B	𝍸 //	7	28
O	𝍸 ////	9	36
AB	////	4	16
		Total 25	100

Grouped Frequency Distributions

When the range of the data is large, the data must be grouped into classes that are more than one unit in width, in what is called a **grouped frequency distribution.** For example, a distribution of the number of hours that boat batteries lasted is the following.

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

Note1: If the data are in tenths, such as 6.2, 7.8, and 12.6, the limits for a class hypothetically might be 7.8–8.8, and the boundaries for that class would be 7.75–8.85. Find these values by subtracting 0.05 from 7.8 and adding 0.05 to 8.8.

Note2: The researcher must decide how many classes to use and the width of each class. To construct a frequency distribution.

Example 2 : These data represent the record high temperatures in degrees Fahrenheit (F) foreach 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

in this example we show the procedure for constructing a grouped frequency distribution, i.e., when the classes contain more than one data value.

1- Find the highest value and lowest value: H =134 and L = 100.

2- Find the range: R = highest value - lowest value = H - L, so

R = 134 - 100 = 34

3- Select the number of classes desired (usually between 5 and 20). In this case, 7 is arbitrarily chosen.

4- Find the class width by dividing the range by the number of classes.

$$\text{Width} = \frac{R}{\text{number of classes}} = \frac{34}{7} = 4.9$$

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
			$n = \Sigma f = 50$

Sometimes it is necessary to use a cumulative frequency distribution. A cumulative frequency distribution is a distribution that shows the number of data values less than or equal to a specific value (usually an upper boundary). As shown below

	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

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 and the boundaries.

Step 2 Tally the data.

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

4.2. Histograms, Frequency Polygons, and Ogives

After you have organized the data into a frequency distribution, you can present them in graphical form. The purpose of graphs in statistics is to convey the data to the viewers in pictorial form. It is easier for most people to comprehend the meaning of data presented graphically than data presented numerically in tables or frequency distributions. This is especially true if the users have little or no statistical knowledge. Statistical graphs can be used to describe the data set or to analyze it. Graphs are also useful in getting the audience's attention in a publication or a speaking presentation. They can be used to discuss an issue, reinforce a critical point, or summarize a data set. They can also be used to discover a trend or pattern in a situation over a period of time.

The three most commonly used graphs in research are

1. The histogram.

2. The frequency polygon.

3. The cumulative frequency graph, or ogive (pronounced o-jive).

An example of each type of graph is shown in Figure 1. The data for each graph are the distribution of the miles that 20 randomly selected runners ran during a given week.

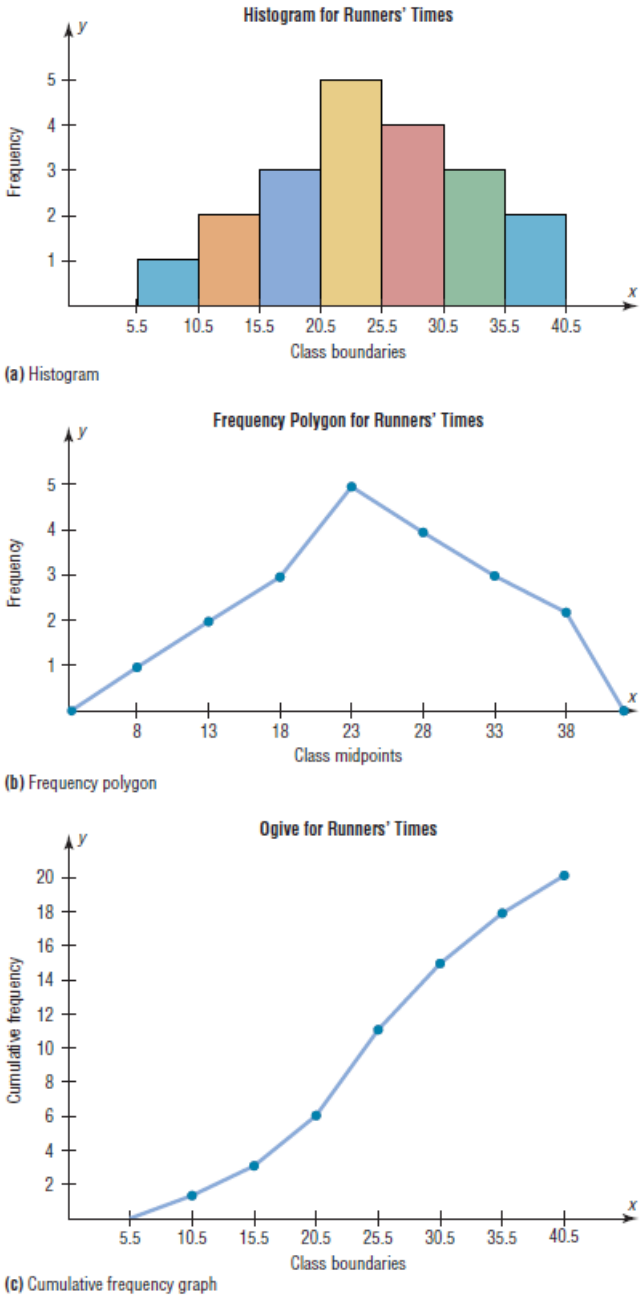


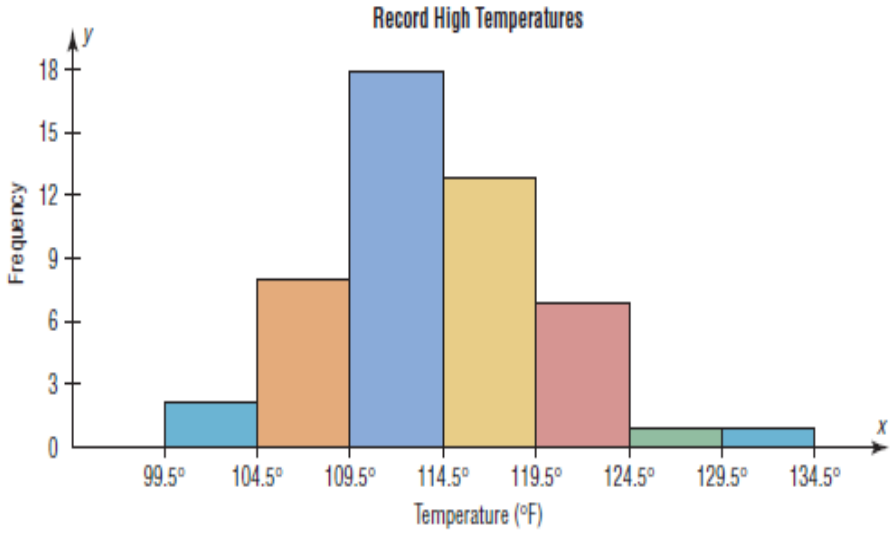
Figure (1) example of Commonly Used Graphs.

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 3: Construct a histogram to represent the data shown for the record high temperatures for each of the 50 states (see Example 2).

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:



The Frequency Polygon

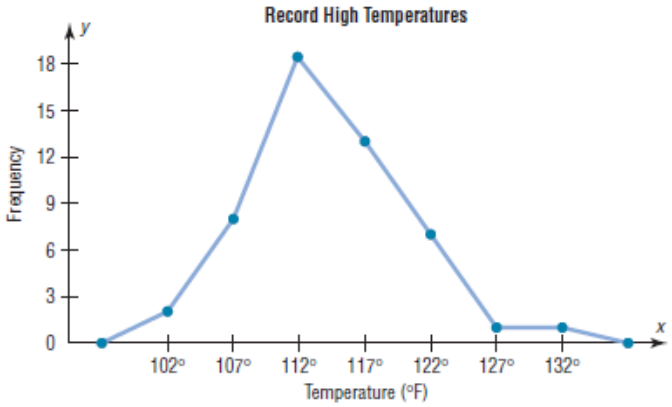
Another way to represent the same data set is by using a frequency polygon

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.

Example 4: Using the frequency distribution given in Example 2, construct a frequency polygon.

Solution: Find the midpoints of each class and draw the relation between frequency and midpoint as shown below

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



The frequency polygon and the histogram are two different ways to represent the same data set. The choice of which one to use is left to the discretion of the researcher.

The Ogive

The third type of graph that can be used represents the cumulative frequencies for the classes. 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.

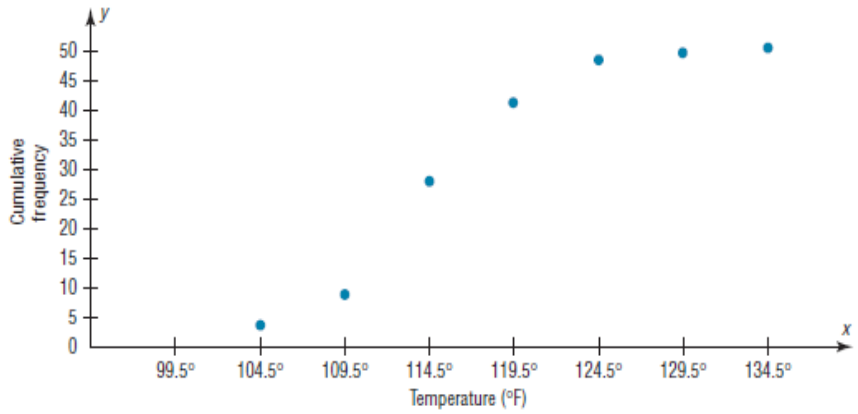
The ogive is a graph that represents the cumulative frequencies for the classes in a frequency distribution.

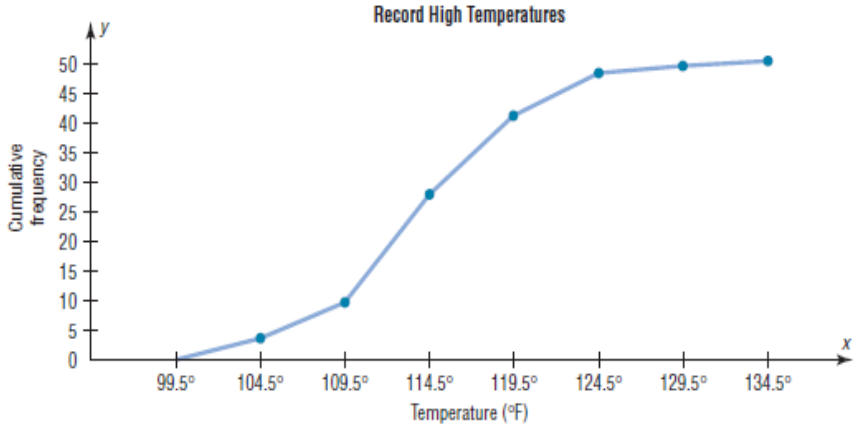
Example 4:

Construct an ogive for the frequency distribution described in Example 2.

	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

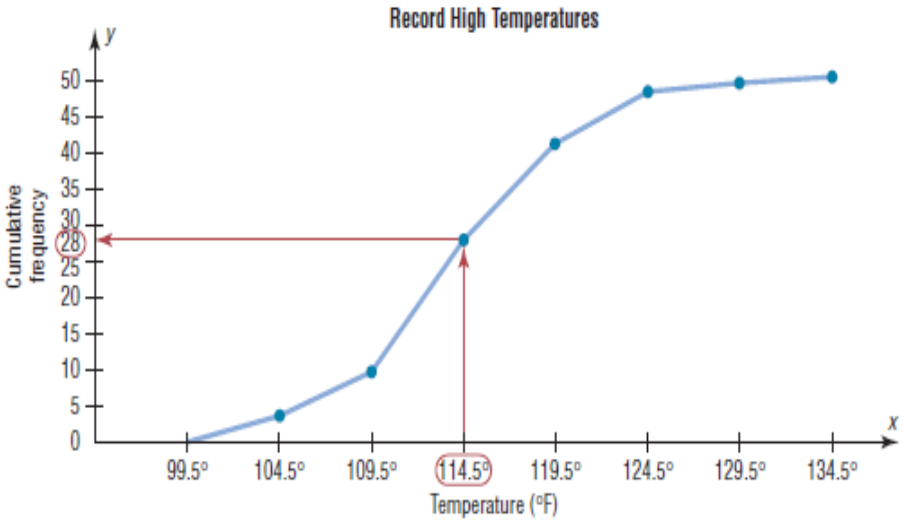
Solution





Cumulative frequency graphs are used to visually represent how many values are below a certain upper-class boundary. For example, to find out how many record high temperatures are less than 114.5 F, locate 114.5_F on the x axis, draw a vertical line up until it intersects the graph, and then draw a horizontal line at that point to the y axis. The y axis value is 28, as shown in Figure below

.



Example 5

Construct a histogram, frequency polygon, and ogive using relative frequencies for the distribution (shown here) of the miles that 20 randomly selected runners ran during a given week.

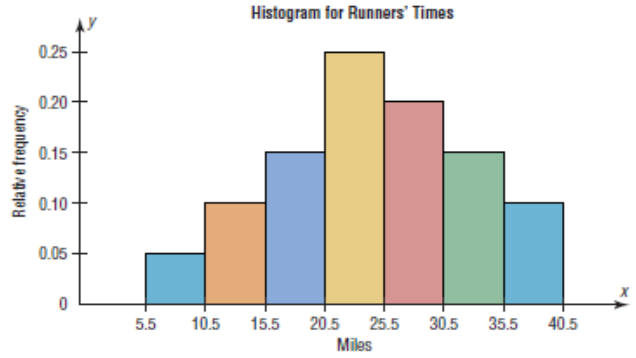
Class boundaries	Frequency
5.5–10.5	1
10.5–15.5	2
15.5–20.5	3
20.5–25.5	5
25.5–30.5	4
30.5–35.5	3
35.5–40.5	2
	20

Solution :

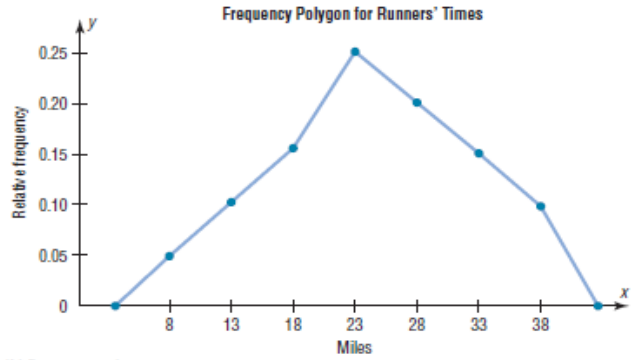
Class boundaries	Midpoints	Relative frequency
5.5–10.5	8	0.05
10.5–15.5	13	0.10
15.5–20.5	18	0.15
20.5–25.5	23	0.25
25.5–30.5	28	0.20
30.5–35.5	33	0.15
35.5–40.5	38	0.10
		1.00

	Cumulative frequency	Cumulative relative frequency
Less than 5.5	0	0.00
Less than 10.5	1	0.05
Less than 15.5	3	0.15
Less than 20.5	6	0.30
Less than 25.5	11	0.55
Less than 30.5	15	0.75
Less than 35.5	18	0.90
Less than 40.5	20	1.00

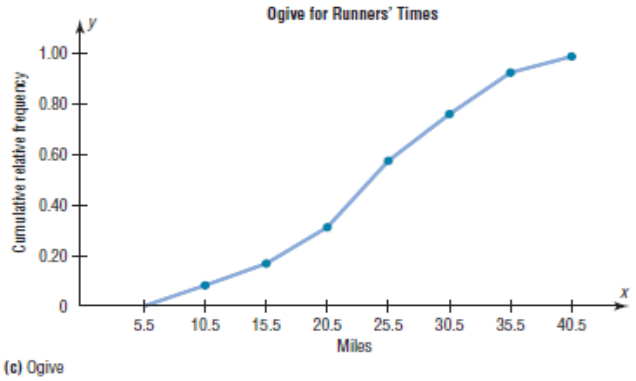
Then , draw the relations as below



(a) Histogram



(b) Frequency polygon



(c) Ogive

Distribution Shapes

When one is describing data, it is important to be able to recognize the shapes of the distribution values. In later chapters you will see that the shape of a distribution also determines the appropriate statistical methods used to analyze the data.

A distribution can have many shapes, and one method of analyzing a distribution is to draw a histogram or frequency polygon for the distribution. Several of the most common shapes are shown in Figure 2:

- the bell-shaped or mound-shaped,
- the uniform shaped,
- the J-shaped and the reverse J-shaped,
- the positively or right-skewed shape,
- the negatively or left-skewed shape,
- the bimodal-shaped, and the U-shaped.

Distributions are most often not perfectly shaped, so it is not necessary to have an exact shape but rather to identify an overall pattern.

A bell-shaped distribution shown in Figure 2(a) has a single peak and tapers off at either end. It is approximately symmetric, i.e., it is roughly the same on both sides of a line running through the center.

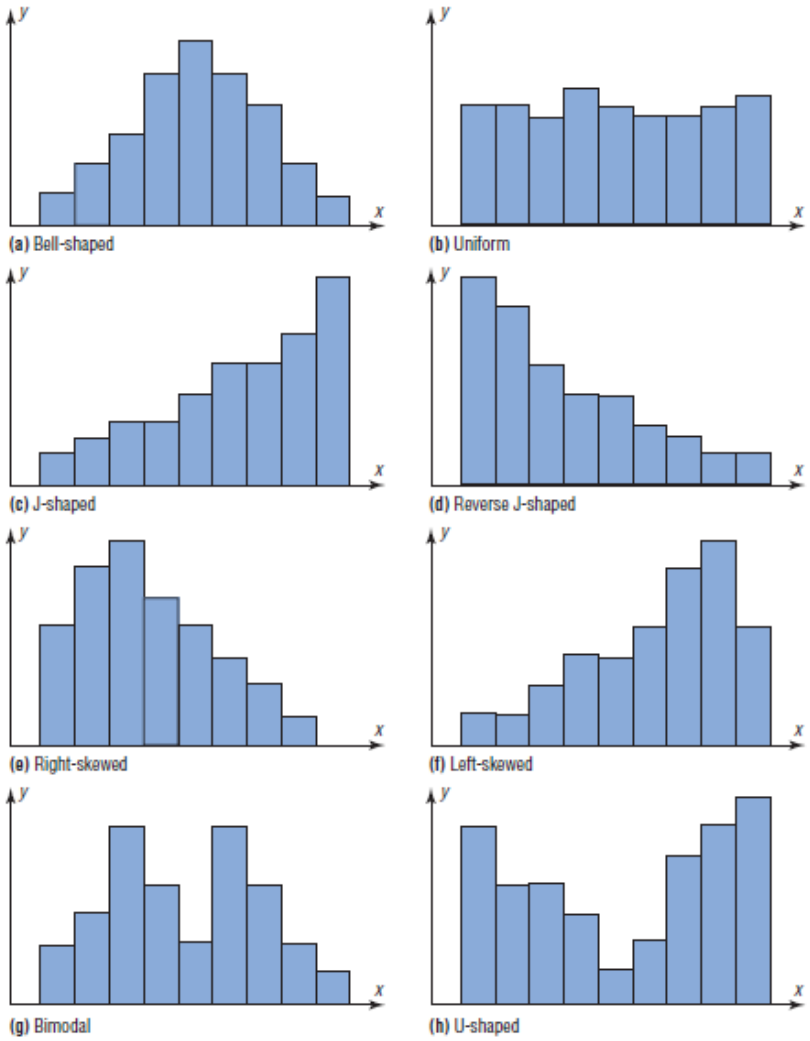


Figure 2 distribution shape

4.3. Other Types of Graphs

In addition to the histogram, the frequency polygon, and the ogive, several other types of graphs are often used in statistics. They are the bar graph, Pareto chart, time series graph, and pie graph. Figure 3 shows an example of each type of graph.

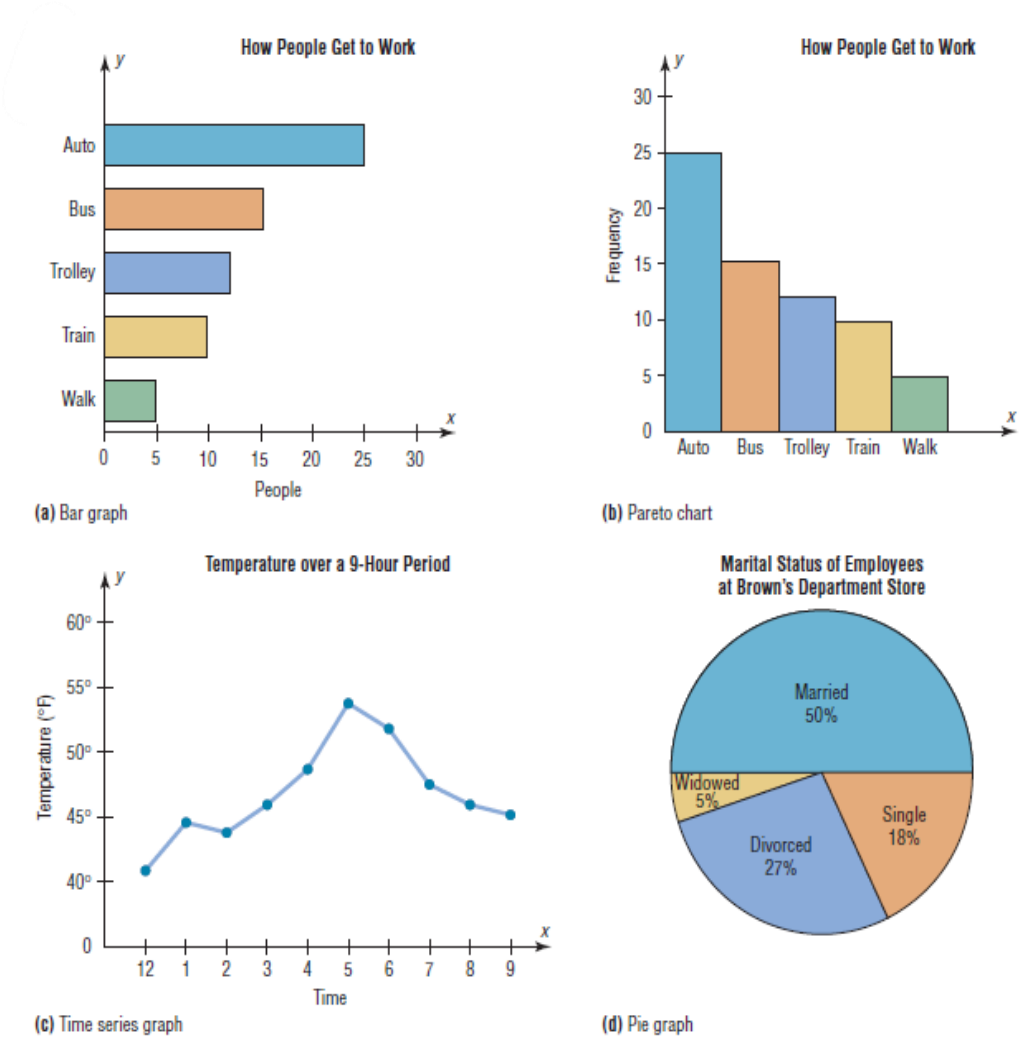


Figure 3. Another type of graphs

Homework 3:

1- The following data represent the ages of some Presidents at the time they were first inaugurated

57	61	57	57	58	57	61	54	68
51	49	64	50	48	65	52	56	46
54	49	50	47	55	55	54	42	51
56	55	54	51	60	62	43	55	56
61	52	69	64	46	54			

- Were the data obtained from a population or a sample? Explain your answer.
- What was the age of the oldest President?

- What was the age of the youngest President?
- Construct a frequency distribution for the data. (Use your own judgment as to the number of classes and class size.)

2- The ages of the Vice Presidents of the United States at the time of their death are listed below. Use the data to construct a frequency distribution, histogram, frequency polygon, and ogive, using relative frequencies. Use 6 classes.

90	83	80	73	70	51	68	79	70	71
72	74	67	54	81	66	62	63	68	57
66	96	78	55	60	66	57	71	60	85
76	98	77	88	78	81	64	66	77	70