Section 1.6 – Additional Displays

Discrete Data

The first step in summarizing quantitative data is to determine whether the data are discrete or continuous. If the data are discrete and there are relatively few different values of the variable, the categories of data (classes) will be the observations (as in qualitative data). If the data are discrete, but there are many different values of the variables, or if the data are continuous, the categories of data (the *classes*) must be created using intervals of numbers.

Construct Histograms of Discrete Data

Definitions

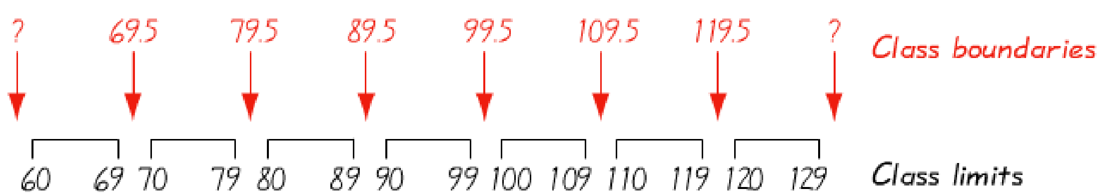
A ***histogram*** is constructed by drawing rectangles for each class of data. The height of each rectangle is the frequency or relative frequency of the class. The width of each rectangle is the same and the rectangles touch each other.

Classes are categories into which data are grouped. When a data set consists of a large number of different discrete data values or when a data set consists of continuous data, we must create classes by using intervals of numbers.

Lower class limits are the smallest numbers that can belong to the different classes (60, 70, 80, 90, 100, 110, and 120) (from previous table).

Upper class limits are the largest numbers that can belong to the different classes (table has upper class limits of 69, 79, 89, 99, 109, 119, 129.)

Class boundaries are the numbers used to separate the classes, but without the gaps created by the class limits. So the complete list of class boundaries is 59.5, 69.5, 79.5, 89.5, 99.5, 109.5, 119.5, 129.5.



Class midpoints are the values in the middle of the classes. From the table: 64.5, 74.5, 84.5, 94.5, 104.5, 114.5, and 124.5.) Each class midpoint is found by adding the lower class limits to the upper class limits and dividing the sum by 2.

Class width is the difference between two consecutive lower class limits or two consecutive lower class boundaries in a frequency distribution. (Table uses a class width of 10.)

Example

The following data represent the time between eruptions (in seconds) for a random sample of 45 eruptions at the Old Faithful Geyser in Wyoming. Construct a frequency and relative frequency distribution of the data.

728	678	723	735	703
730	722	708	714	713
726	716	736	719	672
698	702	738	725	711
721	703	735	699	695
722	718	695	702	731
700	703	706	733	726
720	723	711	696	695
729	699	714	700	718

Solution

The smallest data value is 672 and the largest data value is 738. We will create the classes so that the lower class limit of the first class is 670 and the class width is 10 and obtain the following classes:

Time	*Frequency*	*Relative Frequency*
670 – 679	2	$\frac{2}{45} = 0.044$
680 – 689	0	0
690 – 699	7	$\frac{7}{45} \approx 0.1556$
700 – 709	9	$\frac{9}{45} = 0.2$
710 – 719	9	$\frac{9}{45} = 0.2$
720 – 729	11	$\frac{11}{45} \approx 0.2444$
730 - 739	7	$\frac{7}{45} \approx 0.1556$

- The choices of the lower class limit of the first class and the class width were rather arbitrary.
- There is not one correct frequency distribution for a particular set of data.
- However, some frequency distributions can better illustrate patterns within the data than others. So constructing frequency distributions is somewhat of an art form.
- Use the distribution that seems to provide the best overall summary of the data.

Guidelines for Determining the Lower Class Limit of the First Class and Class Width

Choosing the Lower Class Limit of the First Class.
Choose the smallest observation in the data set or a convenient number slightly lower than the smallest observation in the data set.

Procedure for Constructing a Frequency Distribution

The steps for constructing the frequency distributions are as follows:

1. Determine the number of classes. The number of classes should be between 5 and 20, and the number you select might be affected by the convenience of using round numbers.
2. Calculate the class width.

$$\textit{Class width} \approx \frac{(\text{maximum } \textit{data value}) - (\text{minimum } \textit{data value})}{\textit{number of classes}}$$

   Round this result to get a convenient number (usually round up). If necessary, change the number of classes so that they use convenient values.

3. Choose either the minimum data value or a convenient value below the minimum data value as the first lower class limit.
4. Using the first lower class limit and the class width, list the other lower class limits. (Add the class width to the first lower class limit to get the second lower class limit. Add the class width to the second lower class limit to get the third lower class limit, and so on.)
5. List the lower class limits in a vertical column and then enter the upper class limits.
6. Take each individual data value and put a tally mark in the appropriate class. Add the tally marks to find the total frequency for each class.

Example

Using the pulse rate of females in previous table, follow the above procedure to construct the frequency distribution

Pulse Rates of Females

Pulse Rate	*Frequency*
60 – 69	12
70 – 79	14
80 – 89	11
90 – 99	1
100 – 109	1
110 – 119	0
120 – 129	1

Solution

60 –
70 –
80 –
90 –
100 –
110 –
120 –

***Step* 1**: The number of desired classes is 7.

***Step* 2**: $\textit{Class width} \approx \dfrac{124-60}{7} \approx 9.142857 \approx 10$

***Step* 3**: The minimum data value is 60 and it is also a convenient number.

***Step* 4**: Add the class width of 10 to 60 to get 70 (second lower class limit)
70 + 10 = 80, 80 + 10 = 90, 90 + 10 = 100, 100 + 10 = 110, and 110 + 10 = 120.

***Step* 5**: List the lower class limits vertically as shown in the margin. From this list, we identify the corresponding upper class limits as 69, 79, 89, 99, 109, 119, and 129.

***Step* 6**: Enter a tally mark for each data value in the appropriate class. Then add the tally marks to find the frequencies.

Cumulative Frequency Distribution

The ***cumulative frequency*** for a class is the sum of the frequencies for that class and all previous classes. The ***cumulative frequency distribution*** based on the frequency distribution.

Pulse Rate	*Frequency*	*Cumulative Frequency*
60 – 69	12	12
70 – 79	14	12 + 14 = 26
80 – 89	11	26 + 11 = 37
90 – 99	1	37 + 1 = 38
100 – 109	1	38 + 1 = 39
110 – 119	0	39 + 0 = 39
120 – 129	1	39 + 1 = 40

Dot Plots

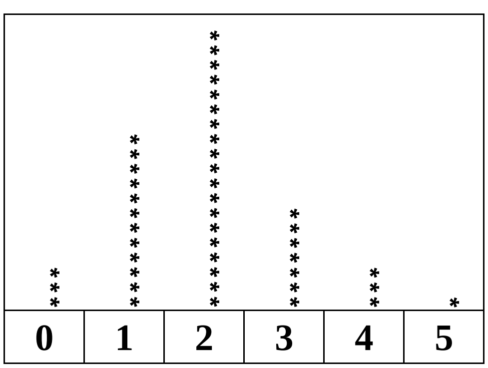
A dotplot consists of a graphin which each data value is plotted as a ***point*** (or ***dot***) along a scale of values. Dots representing equal values are stacked.

Example

The following data represent the number of available cars in a household based on a random sample of 45 households. Draw a dot plot of the data.

3	0	1	2	1	1	2	0	0
4	2	2	2	1	2	4	2	1
1	1	3	2	4	1	2	2	3
3	3	2	1	2	2	2	3	1
2	3	2	1	2	2	5	2	1

Solution



Stemplots

A ***stemplot*** (or ***stem-and-leaf plot***) represents quantitative data by separating each value into two parts: the ***stem*** (such as the *leftmost digit*) and the ***leaf*** (such as the *rightmost digit*)

For ***example***, a data value of 147 would have 14 as the stem and 7 as the leaf.

Example of pulse

Stem (tens)	***Leaves*** (units)	
6	000444488888	← Data values are 60, 60, 60, 64, 64, …, 68
7	2222222266666	← Data values are 72, 72, …
8	00000088888	← Data values are 80, 80, ….
9	6	← Data value is 96
10	4	← Data value is 104
11		
12	4	← Data value is 124

Construction of a Stem-and-leaf Plot

Step 1 The stem of a data value will consist of the digits to the left of the right- most digit. The leaf of a data value will be the rightmost digit.

Step 2 Write the stems in a vertical column in increasing order. Draw a vertical line to the right of the stems.

Step 3 Write each leaf corresponding to the stems to the right of the vertical line.

Step 4 Within each stem, rearrange the leaves in ascending order, title the plot, and include a legend to indicate what the values represent.

Advantage of Stem-and-Leaf Diagrams over Histograms

Once a frequency distribution or histogram of continuous data is created, the raw data is lost (unless reported with the frequency distribution), however, the raw data can be retrieved from the stem-and-leaf plot.

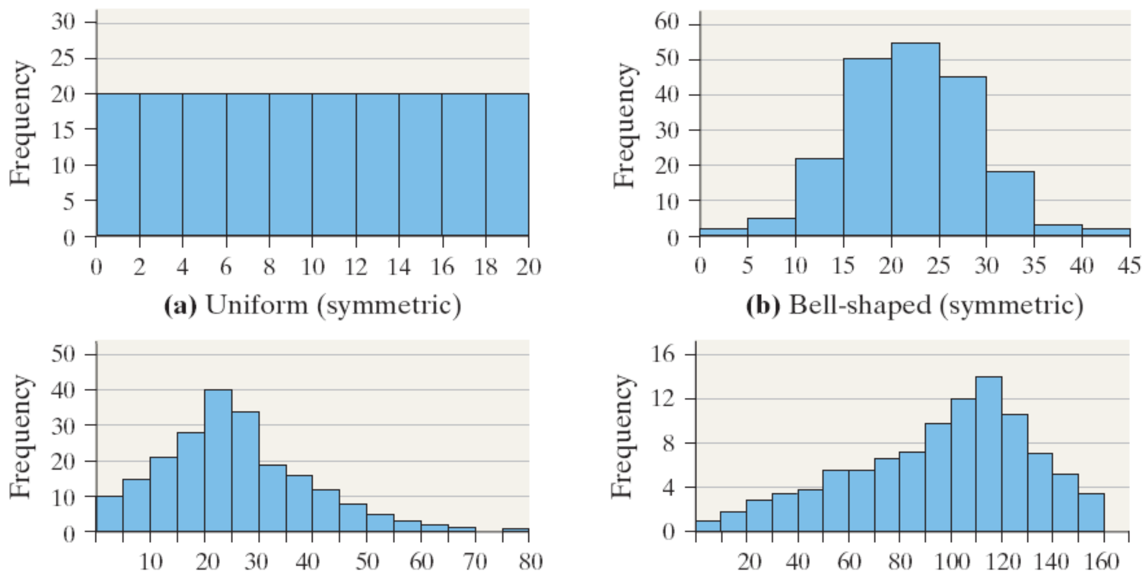
Identify the Shape of a Distribution

Uniform distribution the frequency of each value of the variable is evenly spread out across the values of the variable

Bell-shaped distribution the highest frequency occurs in the middle and frequencies tail off to the left and right of the middle

Skewed right the tail to the right of the peak is longer than the tail to the left of the peak

Skewed left the tail to the left of the peak is longer than the tail to the right of the peak.



(a) Uniform (symmetric)

(b) Bell-shaped (symmetric)

(c) Skewed Right

(d) Skewed Left

Frequency Polygon

A ***class midpoint*** is the sum of consecutive lower class limits divided by 2.

A ***frequency polygon*** is a graph that uses points, connected by line segments, to represent the frequencies for the classes. It is constructed by plotting a point above each class midpoint on a horizontal axis at a height equal to the frequency of the class. Next, line segments are drawn connecting consecutive points. Two additional line segments are drawn connecting each end of the graph with the horizontal axis.

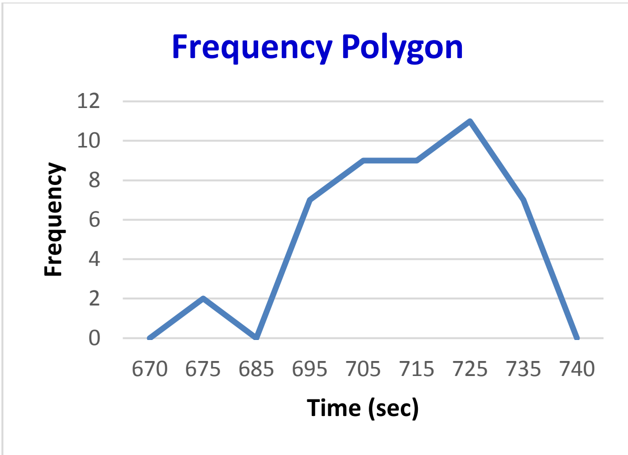
Example

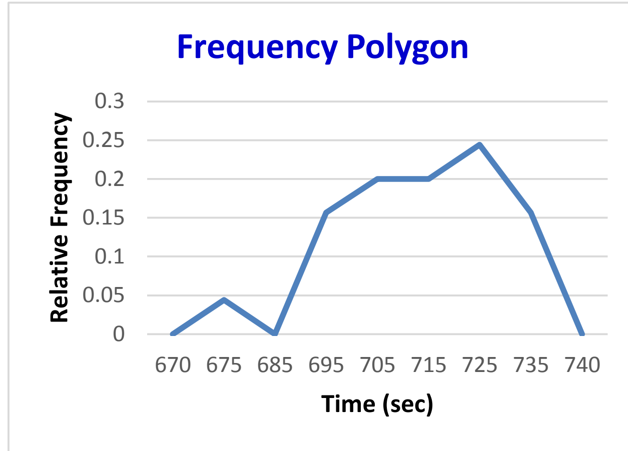
The following data represent the time between eruptions (in seconds) for a random sample of 45 eruptions at the Old Faithful Geyser in Wyoming. Construct a frequency and relative frequency polygon of the data.

728	711	703	678	695	723	735	700	703
730	723	706	722	718	708	714	714	713
726	720	700	716	702	736	719	699	672
698	726	731	702	733	738	725	729	711
721	696	718	703	722	735	699	695	695

Solution

Time	*Class Midpoint*	*Frequency*	*Relative Frequency*
670 – 679	675	2	0.044
680 – 689	685	0	0
690 – 699	695	7	0.1556
700 – 709	705	9	0.2
710 – 719	715	9	0.2
720 – 729	725	11	0.2444
730 - 739	735	7	0.1556





Create Cumulative Frequency and Relative Frequency Tables

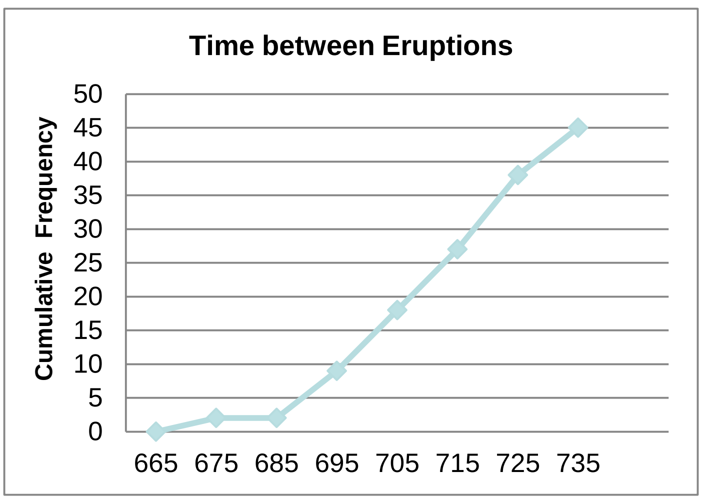
A ***cumulative frequency distribution*** displays the aggregate frequency of the category. In other words, for discrete data, it displays the total number of observations less than or equal to the category. For continuous data, it displays the total number of observations less than or equal to the upper class limit of a class.

A ***cumulative relative frequency distribution*** displays the proportion (or percentage) of observations less than or equal to the category for discrete data and the proportion (or percentage) of observations less than or equal to the upper class limit for continuous data.

Ogive ("**oh-jive**")

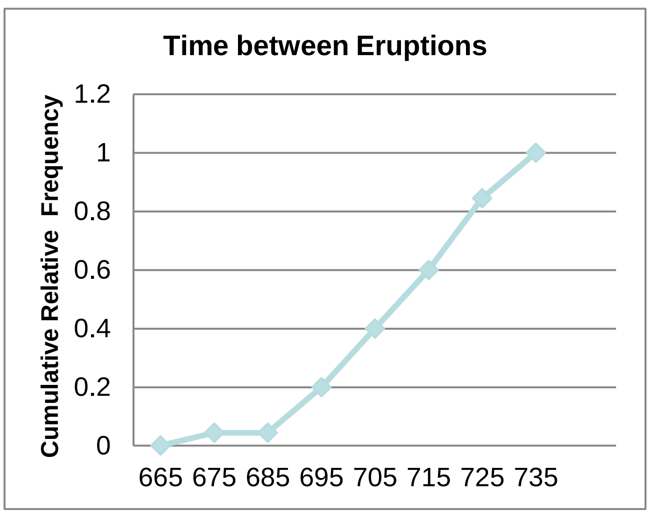
Ogives are useful for determinng the nmber of values below some particular value. An ***ogive*** is a line graph that depicts *cumulative* frequencies. An ogive yses class boundaries along the horizontal scale, and cumulative frequencies along the vertical scale.

Frequency Ogive



Time (seconds)

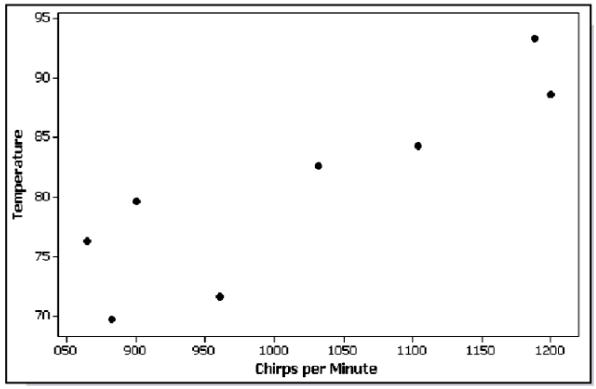
Relative Frequency Ogive



Time (seconds)

Scatter Plot (or Scatter Diagram)

A plot of paired (x,y) data with a horizontal x-axis and a vertical y-axis. Used to determine whether there is a relationship between the two variables



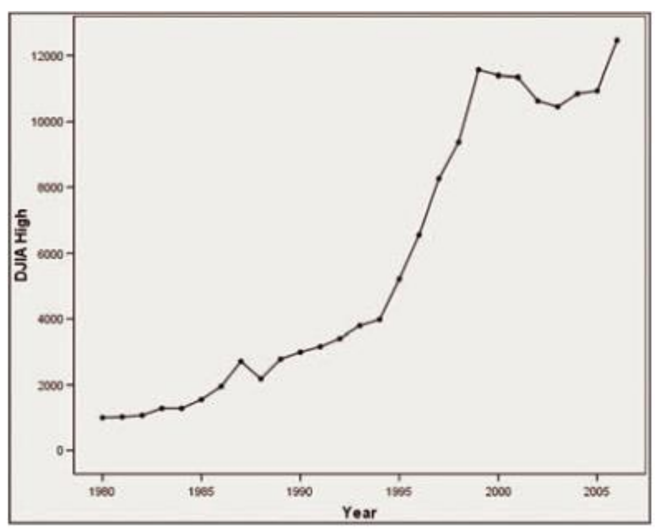
Draw Time Series Graphs

If the value of a variable is measured at different points in time, the data are referred to as ***time series data***.

A ***time-series plot*** is obtained by plotting the time in which a variable is measured on the horizontal axis and the corresponding value of the variable on the vertical axis. Line segments are then drawn connecting the points.

Example

The accompanying SPSS-generated time-series graph shows the yearly high values of the Dow Jones Industrial Average (DJIA) for the N.Y. Stock Exchange. This graph shows a steady increase between the years 1980 and 2007, but the DJIA high values have not been so consistent in more recent years.



Exercise *Section* 1.6 – Additional Displays

1. Identify the class width, class midpoints, and class boundaries for the given frequency distribution. Then construct the cumulative frequency distribution that corresponds to the frequency distribution.

a)

Tar (mg) in Nonfiltered Cigarettes	***Frequency***
10 – 13	1
14 – 17	0
18 – 21	15
22 – 25	7
26 – 29	2

b)

Tar (mg) in Filtered Cigarettes	***Frequency***
2 – 5	2
6 – 9	2
10 – 13	6
14 – 17	15

c)

Weights (lb) of Discarded Metal	***Frequency***
0.00 – 0.99	5
1.00 – 1.99	26
2.00 – 2.99	15
3.00 – 3.99	12
4.00 – 4.99	4

d)

Weights (lb) of Discarded Plastic	***Frequency***
0.00 – 0.99	14
1.00 – 1.99	20
2.00 – 2.99	1
3.00 – 3.99	4
4.00 – 4.99	2
5.00 – 5.99	1

2. Given listed amounts of Strontium-90 (in millibecquerels) in a simple random sample of baby teeth.

155	142	149	130	151	163	151	142	156	133	138	161	128	144	172
137	151	166	147	163	145	116	136	158	114	165	169	145	150	150
150	158	151	145	152	140	170	129	188	156					

a) Construct a dot plot of the amounts of Strontium-90. What does the dot plot suggest about the distribution of those amounts?

b) Construct a stemplot of the amounts of Strontium-90. What does the stemplot suggest about the distribution of those amounts?

c) Construct a frequency polygon of the amounts of Strontium-90. For the horizontal axis, use the midpoints of the class intervals in the frequency distribution: 110-119, 120-129, 130-139, …, 180-189.

d) Construct an ogive of the amounts of Strontium-90. For the horizontal axis, use the class boundaries corresponding to the class limits. How many of the amounts are below 150 millibecquerels?

3. Use the 62 weights if discarded plastic listed in Data set below

0.27	1.41	2.19	2.83	2.19	1.81	0.85	3.05	3.42	2.10	2.93	2.44	2.17	1.41	2.00
0.93	2.97	2.04	0.65	2.13	0.63	1.53	4.69	0.15	1.45	2.68	3.53	1.49	2.31	0.92
0.89	0.80	0.72	2.66	4.37	0.92	1.40	1.45	1.68	1.53	1.44	1.44	1.36	0.38	1.74
2.35	2.30	1.14	2.88	2.13	5.28	1.48	3.36	2.83	2.87	2.96	1.61	1.58	1.15	1.28
0.58	0.74													

a) Construct a dot plot of the weights of discarded plastic. What does the dot plot suggest about the distribution of the weights?

b) Construct a stemplot of the weights of discarded plastic. What does the stemplot suggest about the distribution of the weights?

c) Construct a frequency polygon of the weights of discarded plastic. For the horizontal axis, use the midpoints of the class intervals: 0.00-0.99, 1.00-1.99, 2.00-2.99, 3.00-3.99, 4.00-4.99, 5.00-5.99.

d) Construct an ogive of the weights of discarded plastic. For the horizontal axis, use these classes boundaries: –0.005, 0.995, 1.995, 2.995, 3.995, 4.995, 5.995. How many of the weights are below 4 lb.?

4. In 1965, Intel cofounder Gordon Moore proposed what has since become known as Moore's law: the number of transistors per square inch on integrated circuits with double approximately every 18 months. The table below lists the number of transistors per square inch (in thousands) for several different years. Construct a time-series graph of the data.

Year	1971	1974	1978	1982	1985	1989	1993	1997	1999	2000	2002	2003
Transistors	2.3	5	29	120	275	1180	3100	7500	24,000	42,000	220,000	410,000

5. The following table shows the numbers of cell phone subscriptions (in thousands) in the U.S. for various years. Construct a time-series graph of the data. "Linear" growth would result in a graph that is approximately a straight line. Does the time-series graph appear to show linear growth?

Year	1985	1987	1989	1991	1993	1995	1997	1999	2001	2003	2005
Number	340	1231	3509	7557	16,009	33,786	55,312	86,047	128,375	158,722	207,900

6. The following table lists the marriage and divorce rates per 1000 people in the U.S. for selected years since 1900 (based on data from the Department of Health and Human Services). Construct a multiple bar graph of the data. Why do these data consist of marriage and divorce rates rather than total numbers of marriages and divorces? Comment on any trends that you observe in these rates, and give explanations for these trends.

Year	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
Marriage	9.3	10.3	12.0	9.2	12.1	11.1	8.5	10.6	10.6	9.8	8.3
Divorce	0.7	0.9	1.6	1.6	2.0	2.6	2.2	3.5	5.2	4.7	4.2

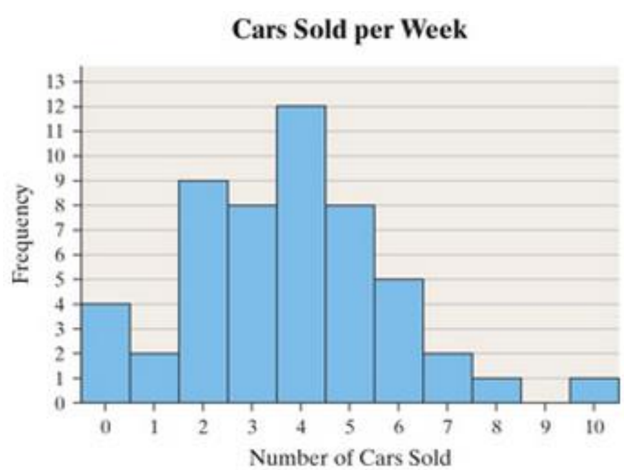
7. A car salesman records the number of cars he sold each week for the past year. The following frequency histogram shows the results

a) What are the most frequent number of cars sold in a week?

b) For how many weeks two cars sold?

c) Determine the percentage of time two cars were sold.

d) Describe the shape of the distribution



8. Use the data to create a stemplot

The midterm test scores for the seventh-period typing class are listed below

85 77 93 91 74 65 68 97 88 59 74 83 85 72 63 79

9. Use the data to create a stemplot. Twenty-four workers were surveyed about how long it takes them to travel to work each day. The data below are given in minutes

20 35 42 52 65 20 60 49 24 37 23 24

22 20 41 25 28 27 50 47 58 30 32 48

10. Find the original data from the stemplot

a)

Stem	*Leaves*
76	2 6 7
77	2 4 9
78	1 7

b)

1	0 1 4
2	1 4 4 7 9
3	3 5 5 5 7 7 8
4	0 0 1 2 6 6 8 9 9
5	3 3 5 8
6	2

c)

24	0 4 7
25	0 2 3 9 9
26	3 4 5 8 8 9
27	0 1 1 3 6 6
28	2 3 8