OBJECTIVES

1. Construct frequency distributions for qualitative data
2. Construct bar graphs
3. Construct pie charts

OBJECTIVE 1
CONSTRUCT FREQUENCY DISTRIBUTIONS FOR QUALITATIVE DATA

The frequency of a category is the number of times it occurs in the data set. A frequency distribution is a table that presents the frequency for each category.

EXAMPLE: Suppose a retailer accepts four types of credit cards and lists the types used by the last 50 customers. Construct a frequency distribution for each type of credit card.

Discover	Visa	Visa	Am. Express	Visa
Visa	Visa	Am. Express	MasterCard	Visa
Am. Express	MasterCard	Visa	Visa	Visa
Visa	Am. Express	Am. Express	MasterCard	Visa
MasterCard	Visa	Discover	Am. Express	Discover
Visa	Am. Express	Discover	Visa	MasterCard
Visa	Visa	Visa	Visa	MasterCard
MasterCard	Am. Express	Visa	MasterCard	Visa
MasterCard	Discover	MasterCard	Visa	Visa
MasterCard	Discover	Am. Express	Discover	Visa

SOLUTION:

Type	Frequency
MasterCard	11
Visa	23
American Express	9
Discover	7

RELATIVE FREQUENCY

A frequency distribution displays how many observations are in each category. Sometimes, we are interested in the proportion of observations in each category. The proportion of observations in a category is called the relative frequency of the category. The relative frequency of a category is the frequency of the category divided by the sum of all frequencies.

Relative Frequency = $\dfrac{\text{frequency}}{\text{sum of all frequencies}}$

EXAMPLE: Construct a relative frequency distribution for the credit card data.

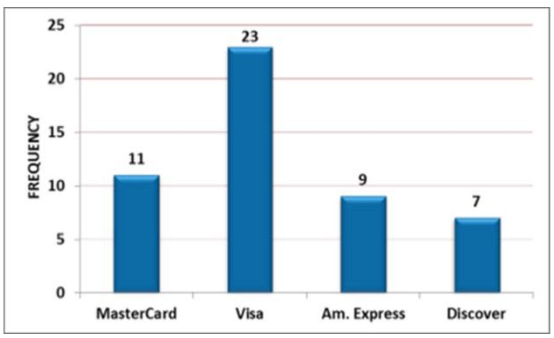
SOLUTION:

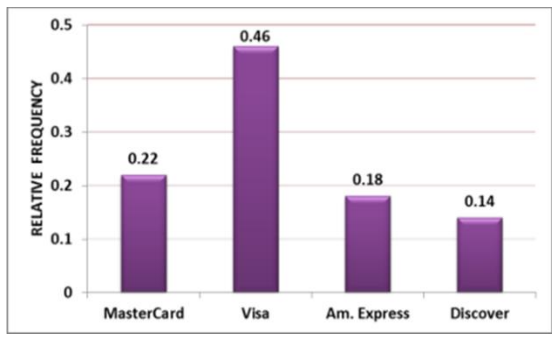
Type	Frequency	Relative Frequency
MasterCard	11	11/50 = 0.22
Visa	23	23/50 = 0.46
American Express	9	9/50 = 0.18
Discover	7	7/50 = 0.14

OBJECTIVE 2
CONSTRUCT BAR GRAPHS

stat → list

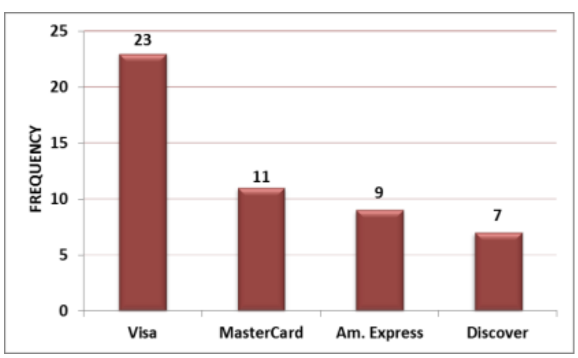
A bar graph is a graphical representation of a frequency distribution. A bar graph consists of rectangles of equal width, with one rectangle for each category. The heights of the rectangles represent the frequencies or relative frequencies of the categories. Following are the frequency and relative frequency bar graphs for the credit card data.





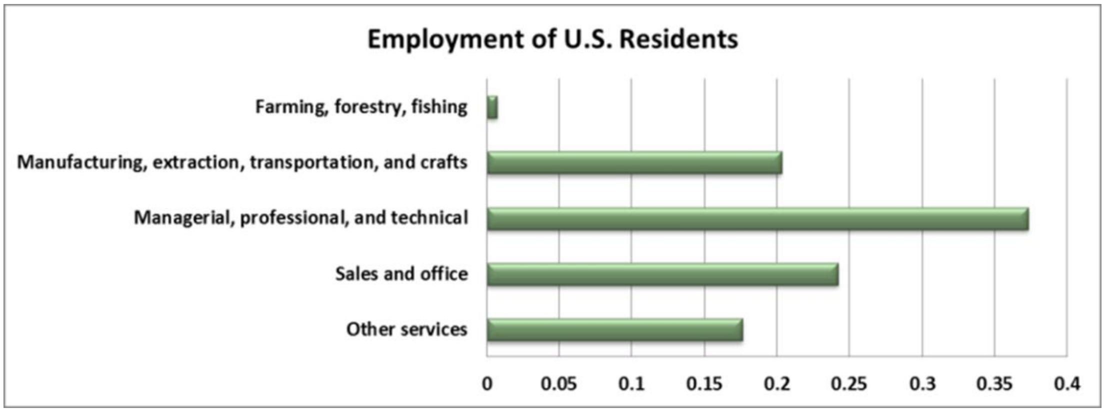
Sometimes it is desirable to construct a bar graph in which the categories are presented in order of frequency or relative frequency. Such a graph is called a pareto chart.

These charts are useful when it is important to see clearly which are the most frequently occurring categories.



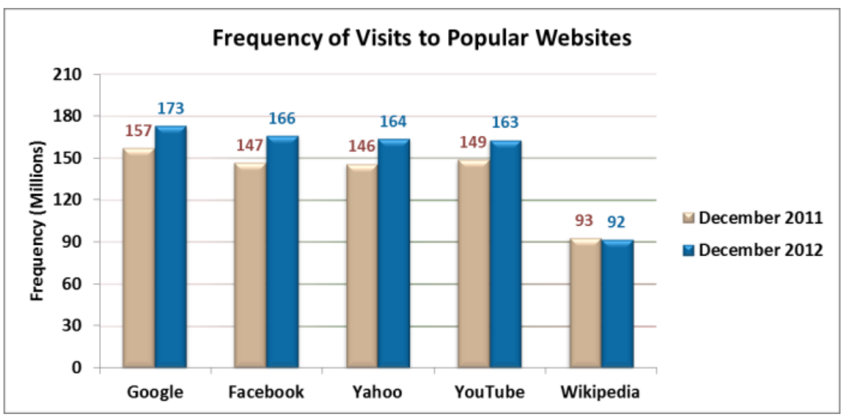
HORIZONTAL BARS

The bars in a bar graph may be either horizontal or vertical. Horizontal bars are sometimes more convenient when the categories have long names.



SIDE-BY-SIDE BAR GRAPHS

Sometimes we want to compare two bar graphs that have the same categories. The best way to do this is to construct both bar graphs on the same axes, putting bars that correspond to the same category next to each other. This is called a side-by-side bar graph.

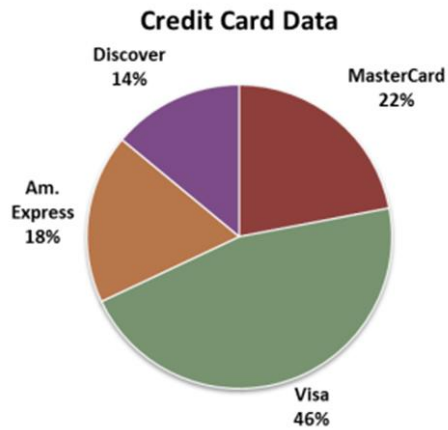


OBJECTIVE 3
PIE CHARTS

A __pie chart__ is an alternative to the bar graph for displaying relative frequency information. A pie chart is a circle which is divided into sectors, one for each category. The relative sizes of the sectors match the relative frequencies of the categories.

For example, if a category has a relative frequency of 0.25, then its sector takes up __25 %__ of the circle. Following is the pie chart for the credit card example at the beginning of this section.

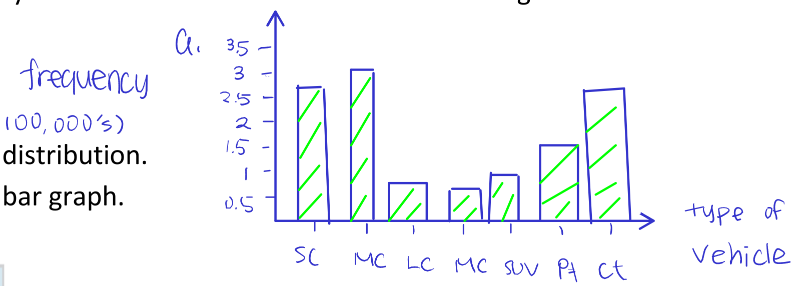
Type	Relative Frequency
MasterCard	0.22
Visa	0.46
Am. Express	0.18
Discover	0.14



Check Your Understanding

The following table presents a frequency distribution for the number of cars and light trucks sold in the month of June 2013.

a. Construct a bar graph.
b. Construct a relative frequency distribution.
c. Construct a relative frequency bar graph.



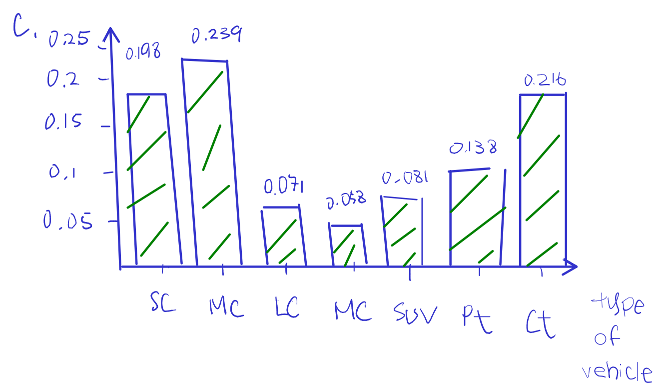
Type of Vehicle	Frequency
Small car SC	276,200
Midsize car mc	333,515
Luxury car LC	98,414
Minivan MV	81,355
SUV SUV	112,328
Pickup truck Pt	191,664
Cross-over truck Ct	300,442

Source: *The Wall Street Journal*

total vehicle = 1393918

b.

type	relative frequency
sc	0.198
mc	0.239
LC	0.071
mv	0.058
SUV	0.081
Pt	0.138
ct	0.216



YOU SHOULD KNOW …

- How to construct a frequency and relative frequency distribution
- How to construct and interpret bar graphs including:
  - Pareto charts
  - Bar graphs with horizontal bars
  - Side-by-side bar graphs
- How to construct and interpret pie charts

OBJECTIVES

1. Construct frequency distributions for quantitative data
2. Construct histograms
3. Determine the shape of a distribution from a histogram
4. Construct frequency polygons and ogives

OBJECTIVE 1
CONSTRUCT FREQUENCY DISTRIBUTIONS FOR QUANTITATIVE DATA

To summarize quantitative data, we use a __frequency distribution__ just like those for qualitative data. However, since these data have no natural categories, we divide the data into __classes__. Classes are intervals of equal width that cover all values that are observed in the data set.

The __lower class limit__ of a class is the smallest value that can appear in that class.

The __upper class limit__ of a class is the largest value that can appear in that class.

The __class width__ is the difference between consecutive lower class limits.

Lower Class Limits

Upper Class Limits

	Class	Frequency
1st	0 – 4	2
2nd	5 – 9	4
3th	10 – 14	9
4th	15 – 19	3

Class Width = 5 – 0 = 5

CHOOSING CLASSES

- Every observation must fall into one of the classes.
- The classes must not overlap.
- The classes must be of equal width.
- There must be no gaps between classes. Even if there are no observations in a class, it must be included in the frequency distribution.

CONSTRUCTING A FREQUENCY DISTRIBUTION

Following are the general steps for constructing a frequency distribution:

Step 1: Choose a class width.

Step 2: Choose a lower class limit for the first class. This should be a convenient number that is slightly less than the minimum data value.

Step 3: Compute the lower limit for the second class, by adding the class width to the lower limit for the first class:
Lower limit for second class = Lower limit for first class + Class width

Step 4: Compute the lower limits for each of the remaining classes, by adding the class width to the lower limit of the preceding class. Stop when the largest data value is included in a class.

Step 5: Count the number of observations in each class, and construct the frequency distribution.

EXAMPLE: The emissions for 65 vehicles, in units of grams of particles per gallon of fuel, are given. Construct a frequency distribution using a class width of 1.

1.5	0.87	1.12	1.25	3.46	1.11	1.12	0.88	1.29	0.94	0.64	1.31	2.49
1.48	1.06	1.11	2.15	0.86	1.81	1.47	1.24	1.63	2.14	6.64	4.04	2.48
1.4	1.37	1.81	1.14	1.63	3.67	0.55	2.67	2.63	3.03	1.23	1.04	1.63
3.12	2.37	2.12	2.68	1.17	3.34	3.79	1.28	2.1	6.55	1.18	3.06	0.48
0.25	0.53	3.36	3.47	2.74	1.88	5.94	4.24	3.52	3.59	3.1	3.33	4.58

SOLUTION: since the smallest value in the data set is 0.25, we choose 0.00 as lower limit for the first class

Class	Frequency
0.00 - 0.99	9
1.00 - 1.99	26
2.00 - 2.99	11
3.00 - 3.99	13
4.00 - 4.99	3
5.00 - 5.99	1
6.00 - 6.99	2

total 65

RELATIVE FREQUENCY

Given a frequency distribution, a relative frequency distribution can be constructed by computing the relative frequency for each class.

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

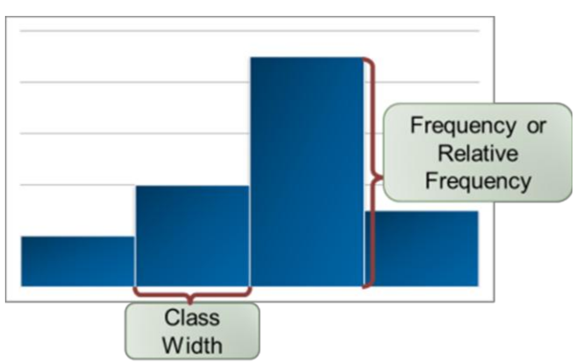
EXAMPLE: Construct a relative frequency distribution for the car emissions data in the last example.

SOLUTION:

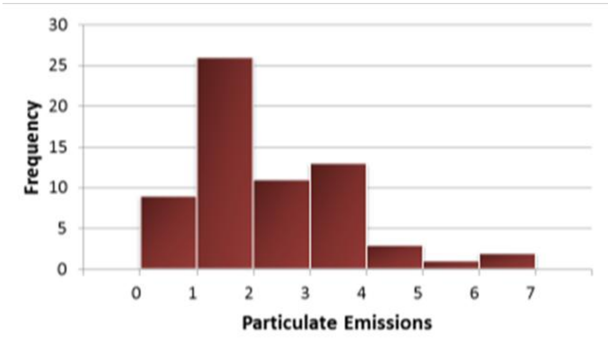
Class	Frequency	Relative Frequency
0.00 – 0.99	9	9/65 = 0.139
1.00 – 1.99	26	26/65 = 0.400
2.00 – 2.99	11	11/65 = 0.169
3.00 – 3.99	13	13/65 = 0.200
4.00 – 4.99	3	3/65 = 0.046
5.00 – 5.99	1	1/65 = 0.015
6.00 – 6.99	2	2/65 = 0.031

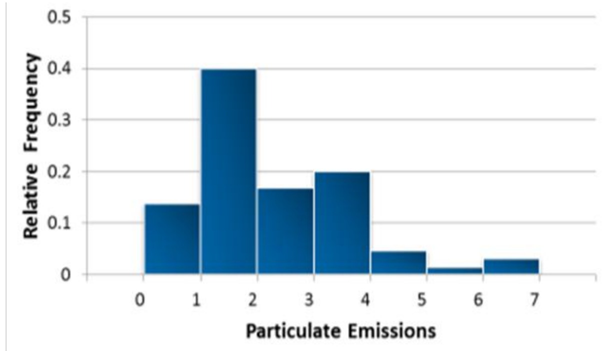
OBJECTIVE 2
CONSTRUCT HISTOGRAMS

Once we have a frequency distribution or a relative frequency distribution, we can put the information in graphical form by constructing a histogram. A histogram is constructed by drawing a rectangle for each class. The heights of the rectangles are equal to the frequencies or the relative frequencies, and the widths are equal to the class width.



EXAMPLE: The frequency histogram and relative frequency histogram are given for the particulate emissions data.



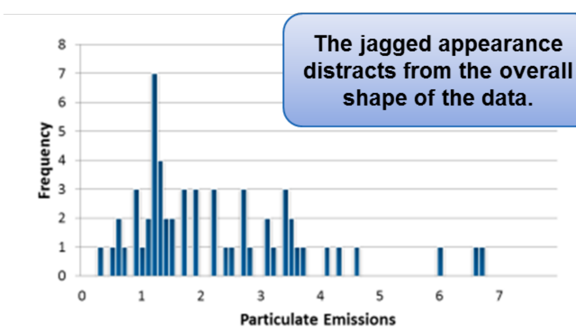


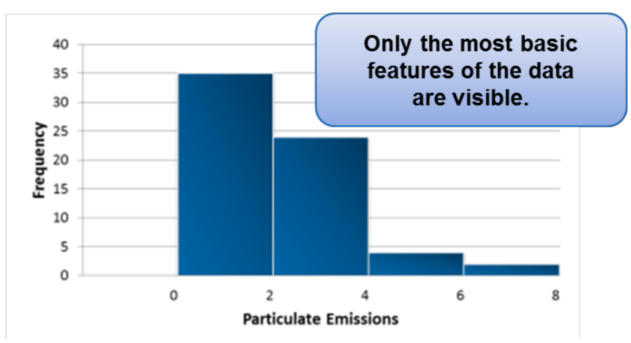
Note that the two histograms have the same shape. The only difference is the scale on the vertical axis.

CHOOSING THE NUMBER OF CLASSES

There are no hard and fast rules for choosing the number of classes. In general, it is good to have more classes rather than fewer, but it is also good to have reasonably large frequencies in some of the classes. There are two principles that can guide the choice:

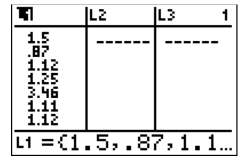
- Too few classes produce a histogram lacking in detail.
- Too many classes produce a histogram with too much detail, so that the main features of the data are obscured.



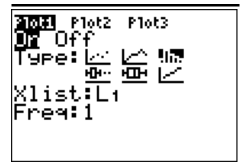


HISTOGRAMS ON THE TI-84 PLUS

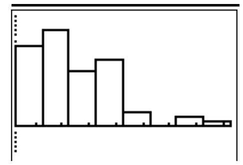
The following steps will create a histogram for the particulate emissions data on the TI-84 PLUS.



Step 1: Enter the data in **L1**.



Step 2: Press **2nd,Y=**, then **1** to access the Plot1 menu. Select **On** and the histogram plot type.



Step 3: Press **Zoom, 9** to view the plot.

OPEN-ENDED CLASSES

It is sometimes necessary for the first class to have no lower limit or for the last class to have no upper limit.

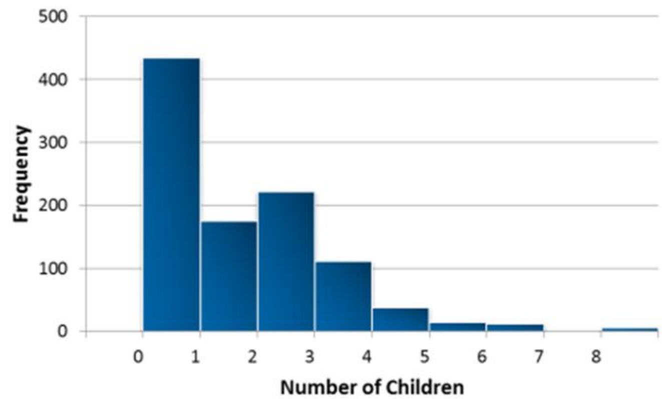
Such a class is called ___open - ended___. The following frequency distribution presents the number of deaths in the U.S. due to pneumonia for various age groups. Note that the last age group is "85 and older", which is an open-ended class. When a frequency distribution contains an open-ended class, a histogram cannot be drawn.

Age	Number of Deaths
5 – 14	69
15 – 24	178
25 – 34	299
35 – 44	875
45 – 54	1872
55 – 64	3099
65 – 74	6283
75 – 84	17,775
85 and older	27,758

HISTOGRAMS FOR DISCRETE DATA

When data are discrete, we can construct a frequency distribution in which each possible value of the variable forms a class. The following table and histogram presents the results of a hypothetical survey in which 1000 adult women were asked how many children they had.

Number of Children	Frequency
0	435
1	175
2	222
3	112
4	38
5	9
6	7
7	0
8	2

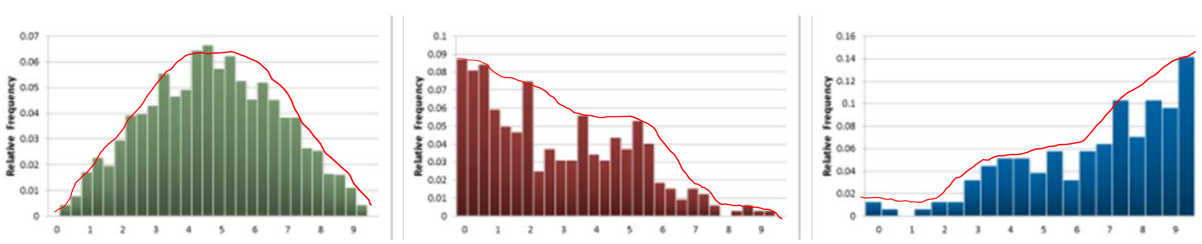


OBJECTIVE 3
DETERMINE THE SHAPE OF A DISTRIBUTION FROM A HISTOGRAM

A histogram gives a visual impression of the "shape" of a data set. Statisticians have developed terminology to describe some of the commonly observed shapes. A histogram is __symmetric__ if its right half is a mirror image of its left half. There are very few histograms that are perfectly symmetric, but many are approximately symmetric. A histogram with a long right-hand tail is said to be

__skewed to the right__.

A histogram with a long left-hand tail is said to be __skewed to the left__.



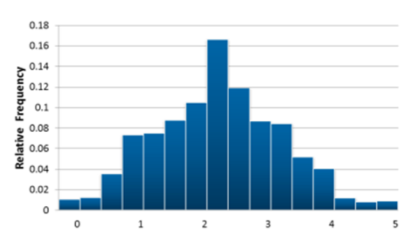
Approximately symmetric

skewed to the right

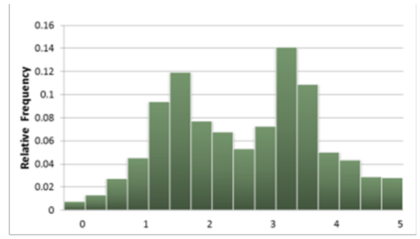
skewed to the left

MODES

A peak, or high point, of a histogram is referred to as a mode. A histogram is ___unimodal___ if it has only one mode, and ___bimodal___ if it has two clearly distinct modes.



Unimodal Histogram



Bimodal Histogram

OBJECTIVE 4
CONSTRUCT FREQUENCY POLYGONS AND OGIVES

CLASS MIDPOINTS

Some graphs used for representing frequency or relative frequency distributions require class midpoints. The midpoint of a class is the average of its lower class limit and the lower class limit of the next class.

Class Midpoint = $\frac{\text{lower limit for a class + lower limit of the next class}}{2}$

EXAMPLE: Consider the classes in the particulate emissions data from earlier in this section find the class midpoints.

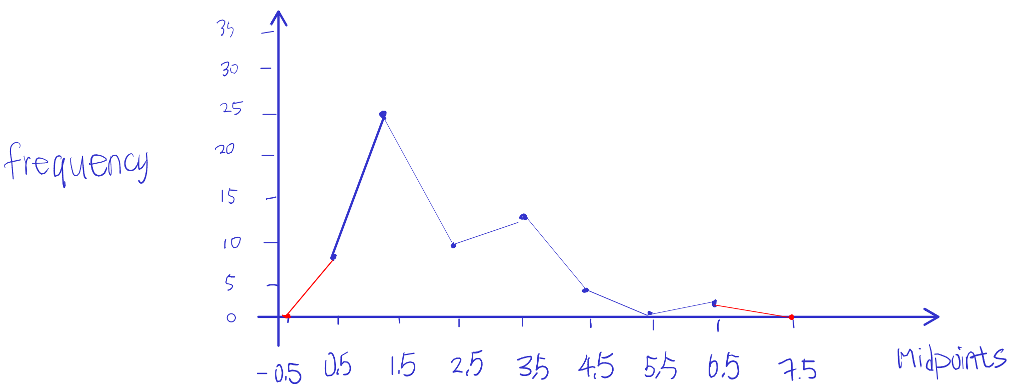
SOLUTION:

Class	Class Midpoint	Frequency
0.00 – 0.99	$\frac{0+1}{2} = 0.5$	9
1.00 – 1.99	$\frac{1+2}{2} = 1.5$	26
2.00 – 2.99	2.5	11
3.00 – 3.99	3.5	13
4.00 – 4.99	4.5	3
5.00 – 5.99	5.5	1
6.00 – 6.99	6.5	2

FREQUENCY POLYGON

Although histograms are the most commonly used graphical display for representing a frequency distribution, there are others. One of these is the frequency polygon. A frequency polygon is constructed by plotting a point for each class. The x coordinate of the point is the class midpoint and the y coordinate is the frequency. Then, all points are connected with straight lines.

EXAMPLE: Consider the classes in the particulate emissions data from earlier in this section. Construct a frequency polygon.



OGIVES AND CUMULATIVE FREQUENCY

Another type of graphical representation of frequency distributions is called an ogive.

Ogives plot values known as cumulative frequency. The cumulative frequency of a class is the sum of the frequencies of that class and all previous classes.

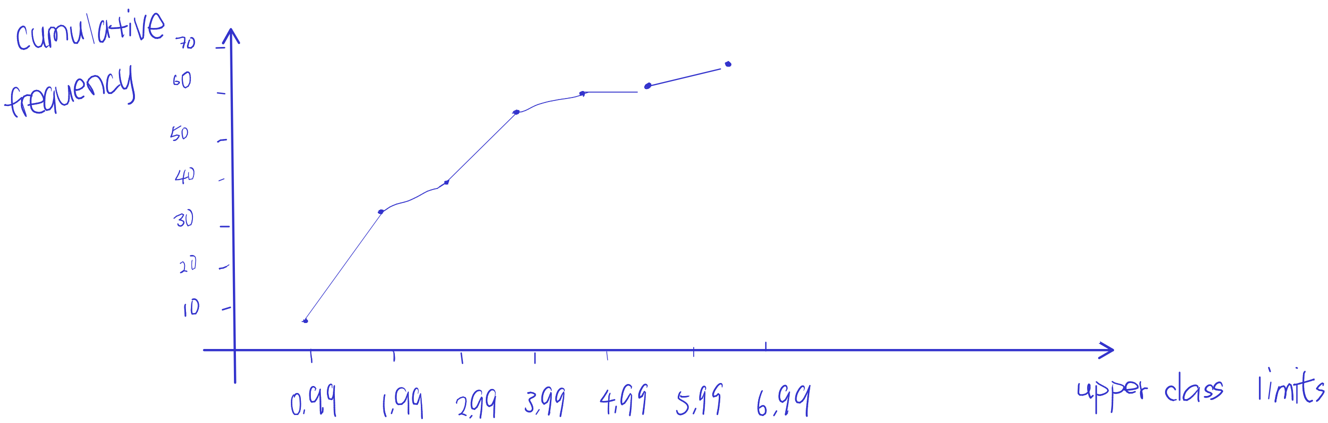
EXAMPLE: Consider the classes in the particulate emissions data from earlier in this section compute the cumulative frequencies.

SOLUTION:

Class	Frequency	Cumulative Frequency
0.00 – 0.99	9	9
1.00 – 1.99	26	25 + 9 = 35
2.00 – 2.99	11	11 + 35 = 46
3.00 – 3.99	13	13 + 46 = 59
4.00 – 4.99	3	= 62
5.00 – 5.99	1	= 63
6.00 – 6.99	2	= 65

upper class limits

An ogive is constructed by plotting a point for each class. The x coordinate of the point is the upper class limit and the y coordinate is the cumulative frequency. Then, all points are connected with straight lines.



YOU SHOULD KNOW …

- How to construct a frequency and relative frequency distribution for quantitative data
- How to construct and interpret histograms
- The guiding principles for choosing the number of classes in a histogram
- How to construct a histogram on the TI-84 PLUS calculator
- Some possible shapes of a data set including:
  - Symmetric
  - Skewed to the right (positively skewed)
  - Skewed to the left (negatively skewed)
  - Unimodal
  - Bimodal
- How to construct and interpret frequency polygons and ogives

OBJECTIVES

1. Construct stem-and-leaf plots
2. Construct dotplots
3. Construct time-series plots

OBJECTIVE 1
CONSTRUCT STEM-AND-LEAF PLOTS

Stem-and-leaf plots are a simple way to display small data sets. In a stem-and-leaf plot, the

__right most__________ digit is the leaf, and the remaining digits form ___stem______.

14 / 8   273 / 9

stem leaf   stem leaf

Consider the values 14.8 and 2,739:

EXAMPLE: The following table presents the U.S. Census Bureau projection for the percentage of the population aged 65 and over for each state and the District of Columbia. Construct a stem-and-leaf plot.

Alabama	14.1	Rhode Island	14.1	Nevada	12.3	Kentucky	13.1
Arkansas	14.3	Tennessee	13.3	New Mexico	14.1	Maryland	12.2
Connecticut	14.4	Vermont	14.3	North Dakota	15.3	Minnesota	12.4
Florida	17.8	West Virginia	16	Oregon	13	Montana	15
Idaho	12	Alaska	8.1	South Carolina	13.6	New Hampshire	12.6
Iowa	14.9	California	11.5	Texas	10.5	New York	13.6
Louisiana	12.6	Delaware	14.1	Virginia	12.4	Ohio	13.7
Massachusetts	13.7	Georgia	10.2	Wisconsin	13.5	Pennsylvania	15.5
Mississippi	12.8	Illinois	12.4	Arizona	13.9	South Dakota	14.6
Nebraska	13.8	Kansas	13.4	Colorado	10.7	Utah	9
New Jersey	13.7	Maine	15.6	D.C.	11.5	Washington	12.2
North Carolina	12.4	Michigan	12.8	Hawaii	14.3	Wyoming	14
Oklahoma	13.8	Missouri	13.9	Indiana	12.7		

SOLUTION:

legend: 8 | 1 = 8.1

Stem	Leaf
8	1
9	0
10	2 5 7
11	5 5
12	0 6 8 4 4 8 3 4 7 2 4 6 2
13	7 8 7 8 3 4 9 0 6 5 9 1 6 7
14	1 3 4 9 1 3 1 1
15	6 3
16	0
17	8

8
9
10
11
12
13
14
15
16
17

BACK-TO-BACK STEM-AND-LEAF PLOTS

When two data sets have values similar enough so that the same stems can be used, their shapes can be compared with a **back-to-back stem-and-leaf plot**.

EXAMPLE: Consider the following course averages from an English class and a History class. The classes can be compared with a back-to-back stem-and-leaf plot.

Course Averages in English Class			
45	57	62	66
68	69	72	74
74	76	77	79
81	83	85	91

Course Averages in History Class			
59	61	67	71
71	73	80	82
82	83	85	88
92	97	98	100

History Class		English Class
	4	5
9	5	7
71	6	2689
311	7	244679
853220	8	135
872	9	1
0	10	

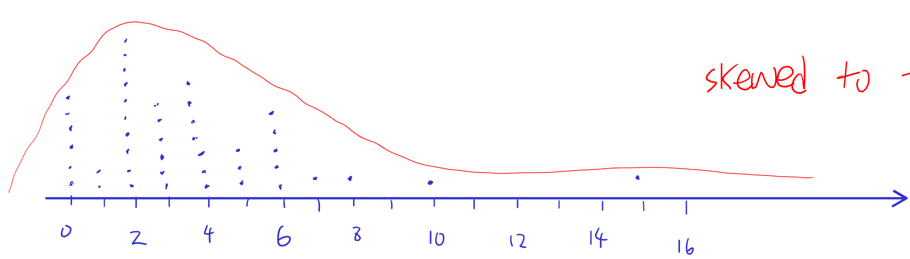
legend: 9 | 5 | 7
59 57

OBJECTIVE 2
CONSTRUCT DOTPLOTS

A dot plot is a graph that can be used to give a rough impression of the shape of a data set. It is useful when the data set is not too large, and when there are some repeated values.

EXAMPLE: Consider the number of children had by each of the presidents of the U.S. and their wives.

0	2	10	2	5	3	6	2	2	4	1	5	4	15	3
4	5	3	2	3	4	2	6	0	0	0	8	3	3	6
2	4	2	0	4	6	4	7	2	0	1	2	6		



skewed to the right

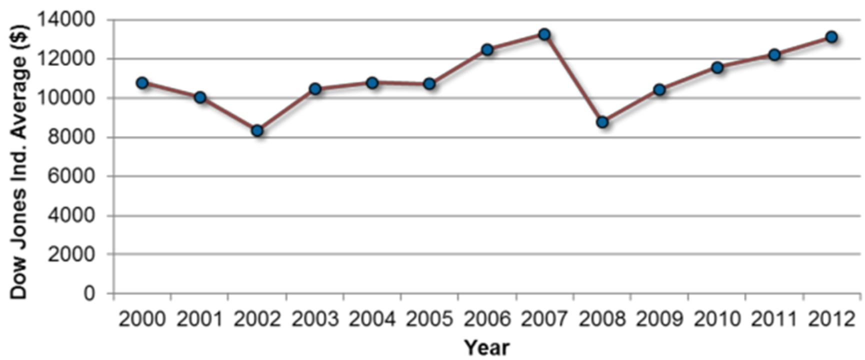
A dotplot gives a good indication of where the values are concentrated, and where the gaps are. For example, it is immediately apparent that the most frequent number of children is 2, and only 4 presidents had more than 6.

OBJECTIVE 3
CONSTRUCT TIME-SERIES PLOTS

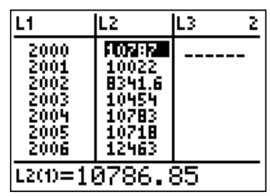
A time - series plot may be used when the data consist of values of a variable measured at different points in time. In a time-series plot, the horizontal axis represents time, and the vertical axis represents the value of the variable we are measuring.

EXAMPLE: The following table and time-series plot display the closing value of the Dow Jones Industrial Average at the end of each year from 2000 to 2012.

Year	Average
2000	10,786.85
2001	10,021.50
2002	8,341.63
2003	10,453.92
2004	10,783.01
2005	10,717.50
2006	12,463.15
2007	13,264.82
2008	8,776.39
2009	10,428.05
2010	11,557.51
2011	12,217.56
2012	13,104.14

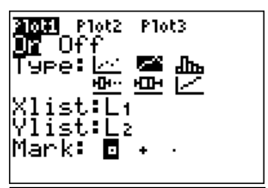


TIME-SERIES PLOTS ON THE TI-84 PLUS

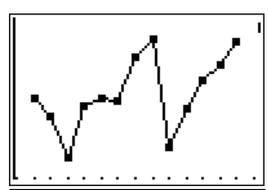


The following steps will create a time-series plot for the Dow Jones Industrial Average data on the TI-84 PLUS.

Step 1: Enter the x-values into **L1** and the y-values into **L2**.



Step 2: Press **2nd**, **Y=**, then **1** to access the Plot 1 menu. Select **On** and the Line Graph plot type. Make sure that **L1** is entered in the **Xlist** field and **L2** is entered in the **Ylist** field.



Step 3: Press **Zoom, 9** to view the plot.

YOU SHOULD KNOW …

- How to construct and interpret:
  - Stem-and-leaf plots
  - Dotplots
  - Time-series plots

OBJECTIVES

1. Understand how improper positioning of the vertical scale can be misleading
2. Understand the area principle for constructing statistical graphs
3. Understand how three-dimensional graphs can be misleading

OBJECTIVE 1
UNDERSTAND HOW IMPROPER POSITIONING OF THE VERTICAL SCALE CAN BE MISLEADING

Statistical graphs, when properly used, are powerful forms of communication. Unfortunately, when graphs are improperly used, they can misrepresent the data and lead people to draw incorrect conclusions. We discuss here three of the most common forms of misrepresentation.

- Incorrect position of the vertical scale
- Incorrect sizing of graphical images
- Misleading perspective for three-dimensional diagrams.

The ___baseline___ of a graph or plot is the value at which the horizontal axis intersects with the vertical axis. With graphs or plots that represent how much or how many of something, it may be misleading if the baseline is not at zero.

These graphs are based on the same data, but give different impressions. The graph on the right has a baseline at ___47___ which exaggerates the differences between the bars.

MISREPRESENTATION WITH TIME-SERIES

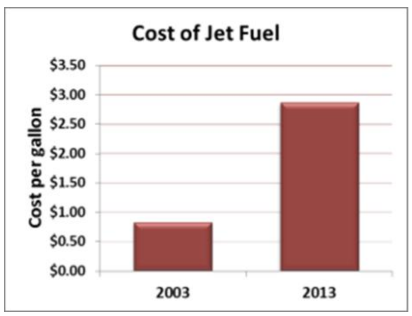
The same misleading information can be created with time-series plots as well.

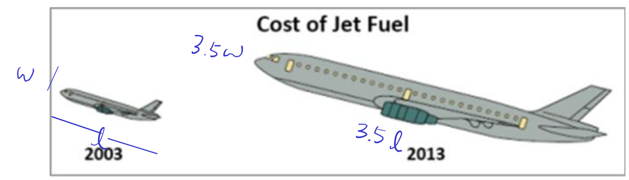
OBJECTIVE 2
UNDERSTAND THE AREA PRINCIPLE FOR CONSTRUCTING STATISTICAL GRAPHS

We often use images to compare amounts. Larger images correspond to greater amounts. To use images properly in this way, we must follow a rule known as **The Area Principle**.

THE AREA PRINCIPLE: when the amounts are compared by constructing an image for each amount, the area of the images must be proportional to amount, for example, if one amount is twice as much as another, its image should be twice as much area as the other image.

EXAMPLE: The cost of jet fuel in 2003 was $0.83 per gallon and in 2013 it had risen to $2.87. Note that the price in 2013 is about 3.5 times the price in 2003.





$A = l \cdot w$

$A = (3.5w)(3.5l)$
$= 12.25 lw$

In the bar graph on the left, the area for the 2013 bar is about 3.5 times that of the 2003 bar. In the picture of the planes, the difference appears much larger. The reason is that both the height and width of the airplane have been increased by a factor of 3.5. Thus, the area of the larger plane is about 12 times the area of the smaller. The airplane graph violates the Area Principle and gives a misleading impression.

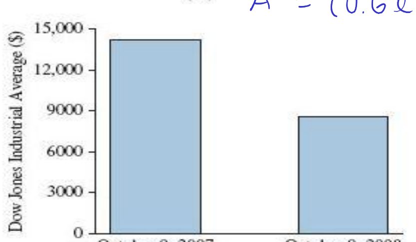
Example: **Stock market crash:** The Dow Jones Industrial Average reached its highest level ever on October 9, 2007, when it closed at $14,164.53. One year later, on October 9, 2008, the average had dropped almost 40%, to $8,579.19. Which of the following graphs accurately represents the magnitude of the drop? Which one exaggerates it?

(A)

$A = lw$

$A = (0.6l)(0.6w) = 0.36lw$

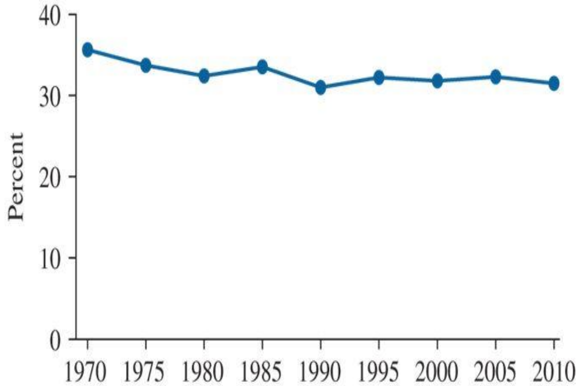
since the smaller bill represents a drop of 64%, instead of a 40% drop, this graph exaggerates the magnitudes of the drop.

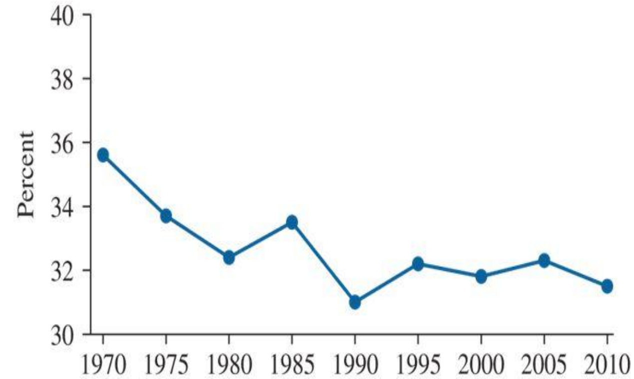


(B)

This graph accurately represents the drop since the bar have the same width and the correct heights, which indicates a 40% drop.

Example College degrees: both of the following time-series plots present the percentage of u.s. bachelor degrees that were in science or mathematics during the years 1970 through 2010. (source: u.s. department of education)





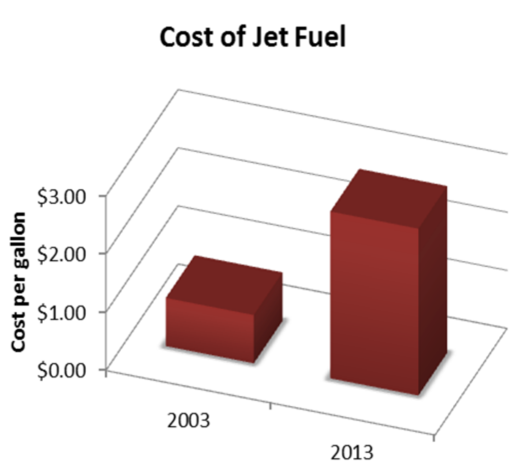
Which of the following statements is more accurate, and why?

- (i) The percentage of degrees that were in science or mathematics decreased considerably between 1970 and 2010.
- (ii) The percentage of degrees that were in science or mathematics decreased somewhat between 1970 and 2010.

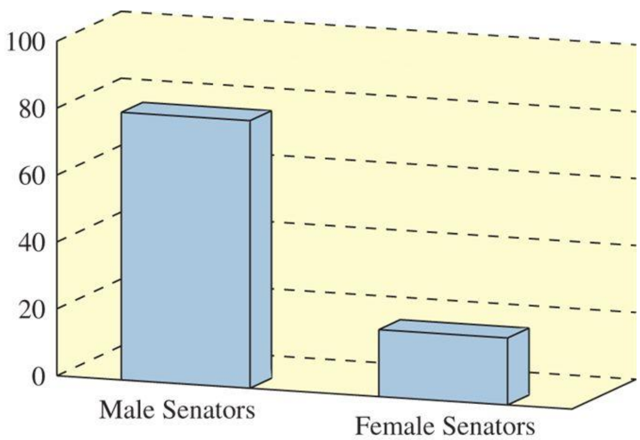
Answer: statement II is more accurate, since statement I is representing the second graph and this graph does not start at zero, this is a misleading graph.

OBJECTIVE 3
UNDERSTAND HOW THREE-DIMENSIONAL GRAPHS CAN BE MISLEADING

Newspapers and magazines often present **three-dimensional** bar graphs because they are visually impressive. Unfortunately, in order to make the tops of the bars visible, these graphs are often drawn as though the reader is looking down on them. This makes the bars look shorter than they really are.



EXAMPLE Female senators: Of the 100 members of the United States Senate in 2013, 80 were men and 20 were women. The following three-dimensional bar graph attempts to present this information.

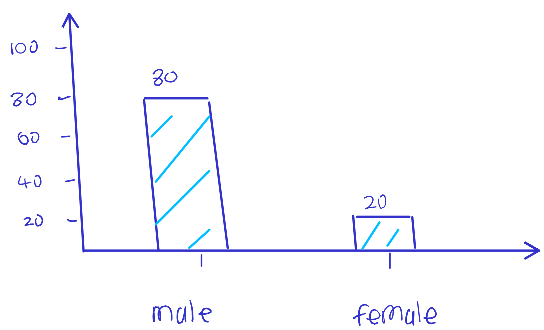


a. Explain how this graph is misleading.
b. Construct a graph (not necessarily three-dimensional) that presents this information accurately.

Answer :

a. The bars for the female and male senators appear shorter than they really are

b.



YOU SHOULD KNOW …

- The common ways that graphs can be misleading including:
  - Incorrect positioning of the vertical axis
  - Incorrect sizing of graphical images by not following the Area Principle
  - How three-dimensional graphs can distort the perspective