Chapter 3

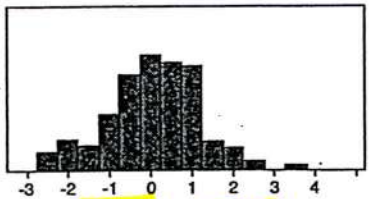
Constructing and Interpreting Graphical Displays of Distributions of Univariate Data

Symmetric graphs appear to have mirror images about their center. If a graph has only one clear peak, it is called **unimodal**; if it has two, it is **bimodal**. Symmetric, unimodal graphs may sometimes be referred to as **mound-shaped**, or **bell-shaped**, because they look like a mound or bell.

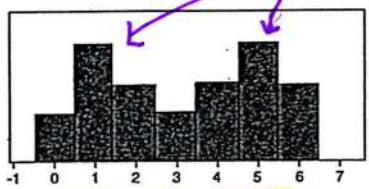
A **uniform** distribution is symmetric where the data are distributed fairly evenly across the graph. There are no clear peaks and the data do not seem to cluster in one area or another.

Below are some examples of graphs and their shapes.

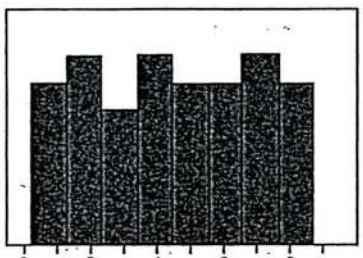
2 PEAKS



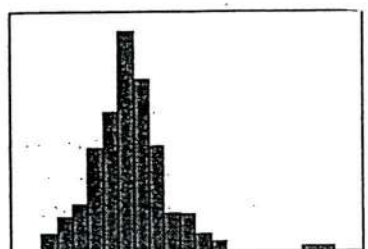
Symmetric and unimodal



Symmetric and bimodal

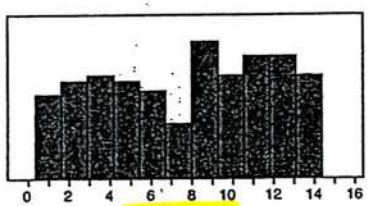


Uniform

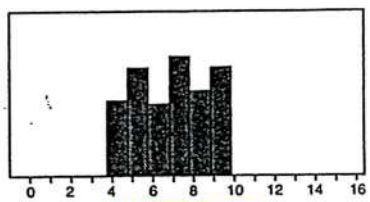


Mound-shaped with outliers

The spread describes the **variability** of the distribution of the data.



Wider Spread



Narrower Spread

Skewed graphs are unimodal graphs that tend to slant—most of the data are clustered on one side of the distribution and "tails" off on the other side. If the tail is on the left, we call the distribution left-skewed. If the tail is on the right, it is right-skewed.

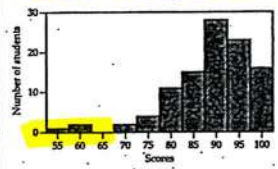
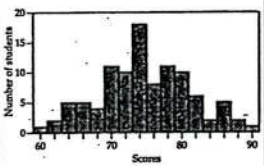
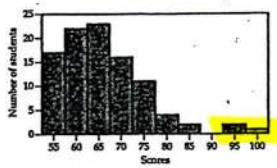
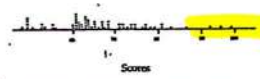
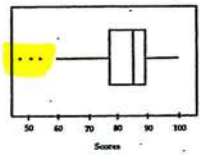
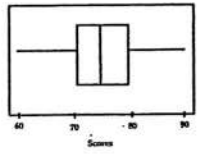
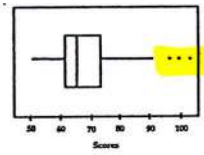
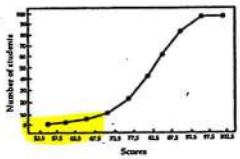
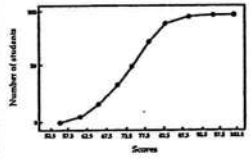
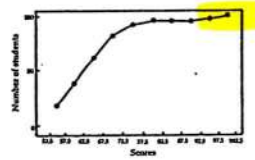
Graph	LEFT-SKEWED OR NEGATIVE SKEW	SYMMETRIC	RIGHT-SKEWED OR POSITIVE SKEW
Histogram			
Dotplot			
Boxplot			
Cumulative Frequency Chart			
Stemplot	4 | 8 5 | 2 5 | 5 6 | 14 6 | 889 7 | 0112233444 7 | 56667778899 8 | 00111111222222333344444444 8 | 55566666677777777788888999999 9 | 222233444 9 | 5555567	6 | 0 6 | 22333 6 | 44555 6 | 667 6 | 88899999 7 | 0000001 7 | 2222222223333333 7 | 4444444444455555 7 | 66677777777 7 | 88899999 8 | 00000 8 | 222222 8 | 445555 8 | 67 8 | 89	5 | 2333344 5 | 55556667799 6 | 00000000011111111122222223333344444 6 | 55555566666666777788899 7 | 0011122224444 8 | 003 8 | 7 9 | 2 9 | 59

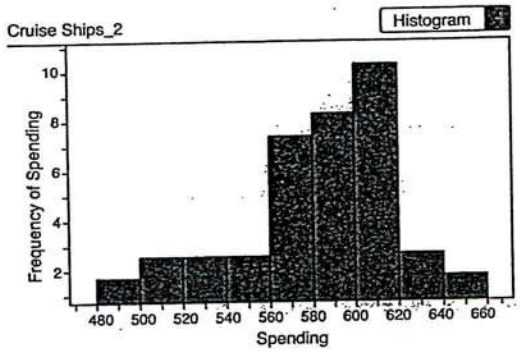
Table 4: Comparison of shapes of different graphs

TEST TIP

Beginning in May 2011, the AP Exam stopped penalizing test takers for incorrect responses to multiple-choice questions. Entering a response for every question—even a wild guess—may help improve your score.

A **histogram** is an appropriate display for quantitative data. It is used primarily for continuous data, but may be used for discrete data that have a wide spread. The horizontal axis is broken into intervals or **bins**. Histograms are also good for large data sets. The histogram below shows the amount of money spent by passengers on a board ship during a recent cruise to Alaska.

EXAMPLE: Describe the distribution below of the amount spent by passengers on board a ship during a recent cruise to Alaska.



11 ANSWER: DISTRIBUTION HAS A SHAPE SKEWED TO THE LEFT. IT PEAKS BETWEEN $600 TO $620. THE DATA IS CENTERED AROUND $590 (MEDIAN). UNIMODAL WITH NO OBVIOUS OUTLIERS.

A **stemplot**, also called stem-and-leaf plot, can be used to display univariate data as well. It is good for small sets of data (about 50 or less) and forms a plot much like a histogram. The stemplot below represents test scores for a class of 32 students.

EXAMPLE: Describe the distribution of test scores for students in the class using the stemplot below.

3	3
4	
5	
6	3 7 9
7	2 2 5 7
8	1 2 6 8 8 8 9 9 9
9	0 0 0 1 3 3 4 5 5 6 7
10	0 0 0 0
Key: 6	3 represents a score of 63

12 ANSWER: THE DISTRIBUTION IS UNIMODAL AND SKEWED TO THE LEFT (OR NEGATIVE). 33% APPEARS TO BE AN OUTLIER. THE DATA IS CENTERED AROUND 89%.

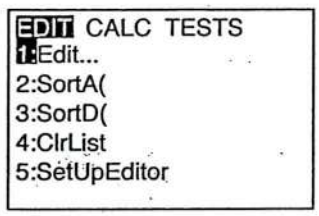
3

[3] DO EXAMPLE AND VERIFY YOUR GRAPH WITH ONE GIVEN

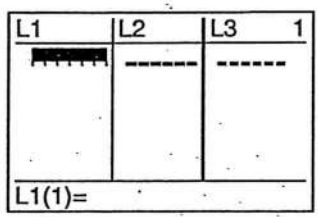
Calculator Tip: INPUT DATA

Many graphs and functions for univariate data can be done on the graphing calculator, but the observations must be first entered into a list. We will use the test scores of the 32 students from a previous example: 33, 63, 67, 69, 72, 72, 75, 77, 81, 82, 86, 88, 88, 88, 89, 89, 89, 90, 90, 90, 91, 93, 93, 94, 95, 95, 96, 97, 100, 100, 100, 100.

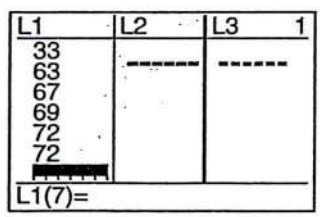
From the home screen, press STAT.



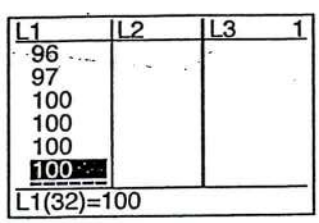
Choose 1:Edit. This brings you into the main list editor.



Enter the data into the list that you are going to use. For now, use L1. Press ENTER after each entry to move to the next line.



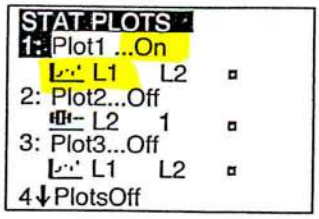
Continue to type in values until the list is complete. Note that the cursor is on the last entry, which in this case is 100, that is the 32nd entry into this list.



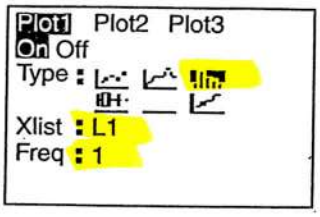
Calculator Tip: CREATE STAT PLOT

3 CONT

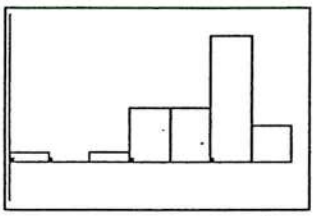
Histograms can be made from data in lists. From the home screen, press 2nd STAT PLOT, and then choose 1:Plot1….



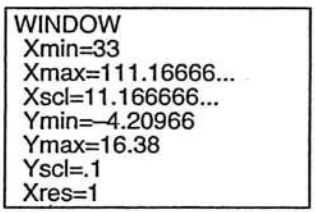
Turn on the plot by pressing ENTER when the cursor is over On. Arrow right to the third type of plot, which is a histogram. Press ENTER so that it is highlighted. Arrow down to Xlist: and press 2nd L1. On the Freq: line, enter 1, if it is not already so. (This is the frequency of each value in Xlist and is used if two lists act as a frequency table.)

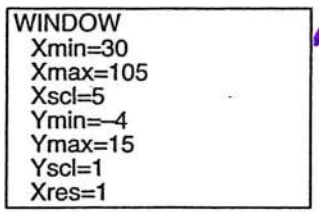


Press ZOOM 9 to see the graph. The calculator will sometimes choose bin widths that are hard to use.

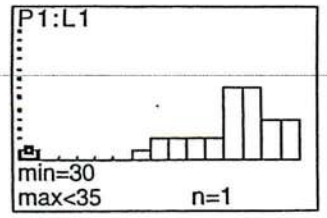


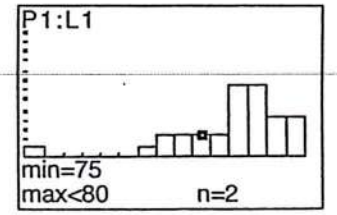
Press WINDOW. A bin width of 5 or 10 may make more sense in this situation than the one of 11.16. Change Xmin, Xmax, Xscl, Ymin, Ymax, and Yscl to the values shown below on the right. Xscl is the width of your bins, beginning at Xmin. An appropriate window for these data is shown on the screen below.





Press GRAPH to view the histogram.





You can use the TRACE button and arrows to view the frequencies of each bin.

SECTION 3.4:

Exploring Bivariate Data

Analyzing Patterns in Scatterplots

Bivariate data consist of two variables, between which one is typically looking for an association. The variables may be categorical or quantitative; in this section we will focus on quantitative bivariate data.

The two variables under study are referred to as the **explanatory variable** (x) and the **response variable** (y). The explanatory variable *explains* or *predicts* the response variable. The response variable measures the outcomes that have been observed.

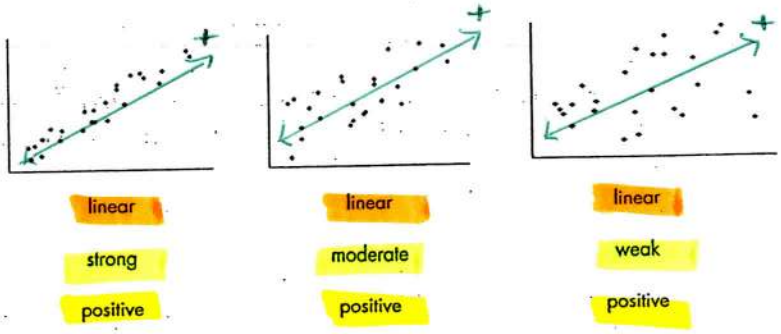
EXAMPLE: Data collected from snack foods included the number of grams of fat per serving and the total number of calories in the food. Identify the explanatory and response variables when looking for a relationship between fat grams and calories.

14 ANSWER: EXPLANATORY VARIABLE: GRAMS OF FAT
RESPONSE VARIABLE: CALORIES
* NUMBER OF GRAMS OF FAT WOULD PREDICT CALORIES IN THE SNACK.

Scatterplots are used to visualize quantitative bivariate data. These plots can tell us if and how two variables are related. When examining univariate data we described a distribution's shape, center, spread, and outliers/unusual features. In a scatterplot, we will focus on its shape, direction, and strength, and look for outliers and unusual features. Below is a scatterplot of the top 30 leading scorers in the National Basketball Association (NBA). Each point represents 1 of the 30 players. Michael Jordan, who scored 32,292 points in 1,072 games, is noted.

The **shape** of a plot is usually classified as linear or nonlinear (curved). The **direction** of a scatterplot tells what happens to the response variable as the explanatory variable increases. This is the slope of the general pattern of the data. The **strength** describes how tight or spread out the points of a scatterplot are.

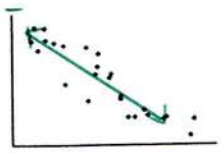
The table below shows comparisons of scatterplots of various shapes, directions, and strengths.



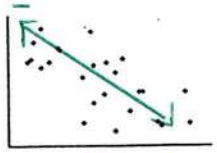
linear
strong
positive

linear
moderate
positive

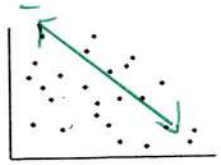
linear
weak
positive



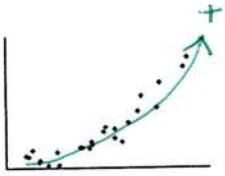
linear
strong
negative



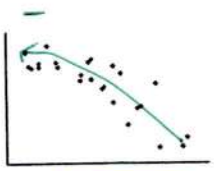
linear
moderate
negative



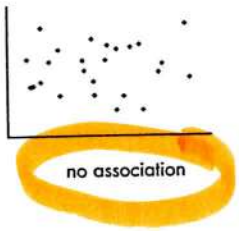
linear
weak
negative



nonlinear
positive
strong

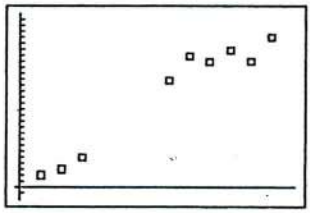


nonlinear
negative
moderate

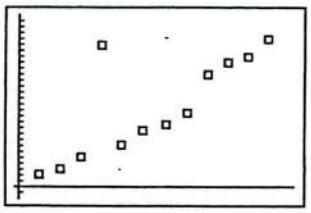


no association

When analyzing a scatterplot it is also a good idea to look for outliers, clusters, or gaps in the data. The scatterplot below has an obvious gap. There is an overall positive, linear association but we should find out the reason for the gap.

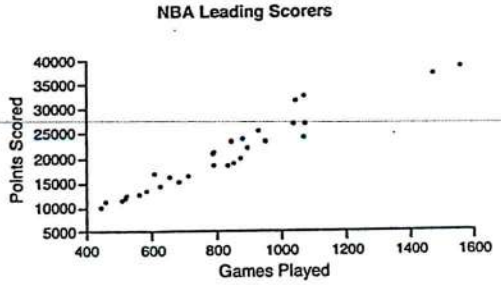


The scatterplot below has an obvious outlier. An outlier falls outside the general pattern of the data. There could be several possible reasons for the outlier and it merits investigation.



15 EXAMPLE: A scatterplot of the top 30 scorers in NBA history is shown below. Identify the explanatory variable and the response variable. Describe the association between the two variables.

ANSWER:



16 TRY THIS SCATTER PLOT

Calculator Tip: INPUT BIVARIATE DATA

A scatterplot can be viewed on the graphing calculator. First, the data must be entered into lists. Recall that you access the list editor by pressing [STAT] and then choosing 1:Edit....

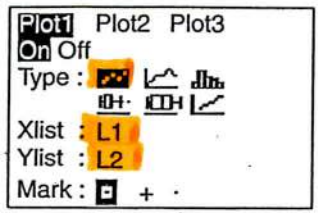
L1	L2	L3 2
2	1	------
3	3	
5	7	
6	7	
9	11	
10	12	

L2(7) =

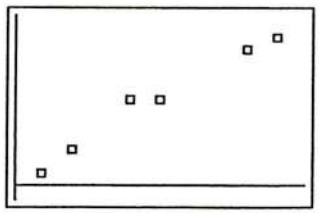
← INPUT DATA

CREATE SCATTER PLOT :

Press [2nd] [STAT PLOT] and choose 1:Plot 1. Turn on the plot, select the scatterplot icon, and enter the appropriate lists for Xlist: and Ylist:.



Press [ZOOM] [9] to see the scatterplot.



← OUTPUT GRAPH