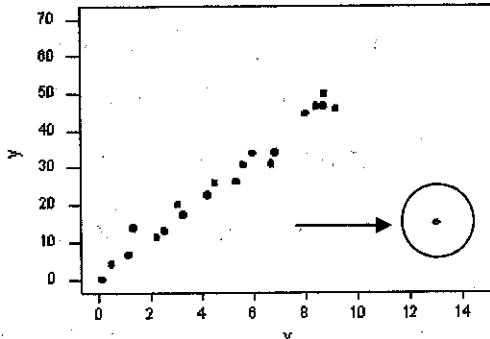
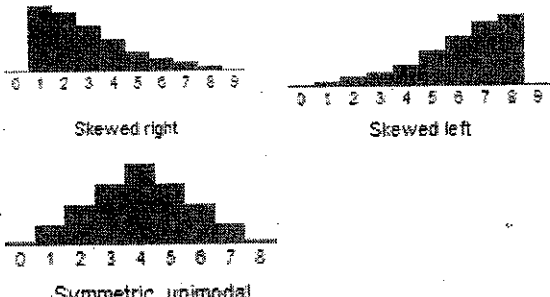


Intro Notes

Term	Describe	Example
Box Plot	Displays distrib. of data based on five # summary Q1, Q3, Med.	<p>Minimum Lower Quartile (Q₁) Median Upper Quartile (Q₃) Maximum</p>
Dot Plot	Display of data using dots (simple scale)	
Histogram	Display of data using bars of different heights * group #s into ranges	
Median	- Middle - Medpt of values	<p>median of all data, second quartile</p> <p>65, 65, 70, 75, 80, 80, 85, 90, 95, 100</p> <p>median of lower part, first quartile median of upper part, third quartile</p>
First and Third Quartiles	1 st - Median of lower half of data 3 rd - Median of upper half of data	<p>median of all data, second quartile</p> <p>65, 65, 70, 75, 80, 80, 85, 90, 95, 100</p> <p>median of lower part, first quartile median of upper part, third quartile</p>

data

Interquartile Range	Measure of variability $Q_3 - Q_1$	Subtract Third Quartile (Q_3) - First Quartile (Q_1) = IQR
Outlier	Data situated away or detached from the main body or system	
Mean	Average	$5 + 4 + 2 + 6 + 3 = 20$ $\frac{20}{5} = 4$ The Mean is 4.
Mean Absolute Deviation (MAD)	Average distance between each data pt and the mean	Steps: 1. Find the Mean 2. Calculate the absolute value of the difference between each data value and the mean 3. Determine the average of the differences in step 2. This average is the mean absolute deviation
Measures of Center	Mean, Median, Mode Middle of data set	Find the Mean and Median for the following data. Hint: (Must order the numbers first before finding the Median) 2 1 5 4 3 Mean: $\frac{15}{5} = 3$ Median = 3
Measures of spread	Describes how similar or varied the data is	Examples of Measures of Spread: 1. Range 2. Interquartile Range (IQR) 3. Mean Absolute Deviation -MAD

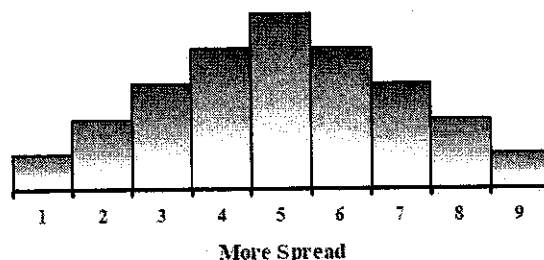
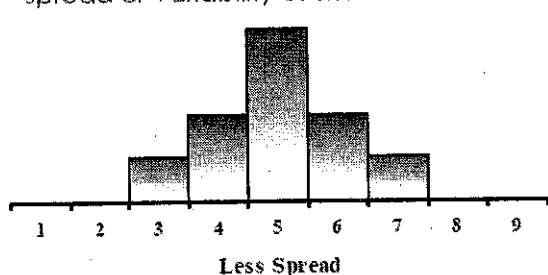
Bivariate Data	<ul style="list-style-type: none"> • Involves 2 variables • Causes/relationships 	Is there a relationship between the number of females in computer programming & their scores in mathematics?
Shape	Shape of the distribution	 <p>The figure shows three histograms. The first, labeled 'Skewed right', has a long tail extending to the right. The second, labeled 'Skewed left', has a long tail extending to the left. The third, labeled 'Symmetric, unimodal', is a bell-shaped curve centered in the middle.</p>

When you compare two or more data sets, focus on four features:

- **Center.** Graphically, the center of a distribution is the point where about half of the observations are on either side.
- **Spread.** The spread of a distribution refers to the variability of the data. If the observations cover a wide range, the spread is larger. If the observations are clustered around a single value, the spread is smaller.
- **Shape.** The shape of a distribution is described by symmetry, skewness, number of peaks, etc.
- **Unusual features.** Unusual features refer to gaps (areas of the distribution where there are no observations) and outliers.

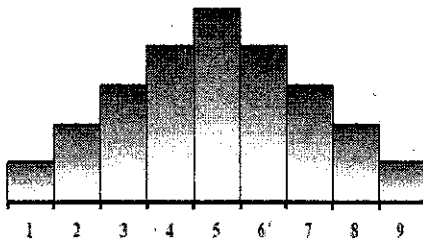
SPREAD

The spread of a distribution refers to the variability of the data. If the data cluster around a single central value, the spread is smaller. The further the observations fall from the center, the greater the spread or variability of the set.

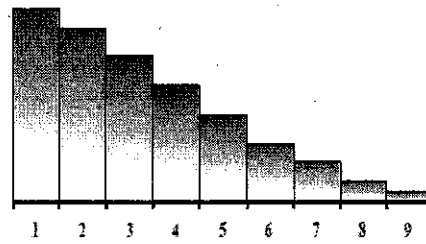


SHAPE

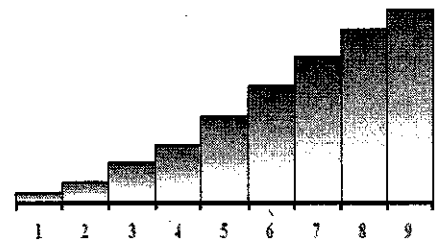
The shape of a distribution is described by symmetry, number of peaks, direction of skew, or uniformity



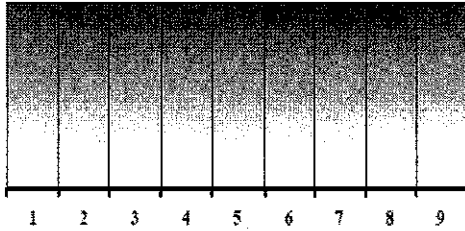
symmetric, unimodal, bell-shaped



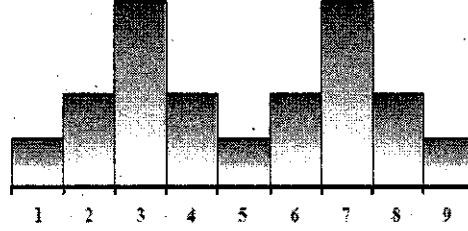
skewed right



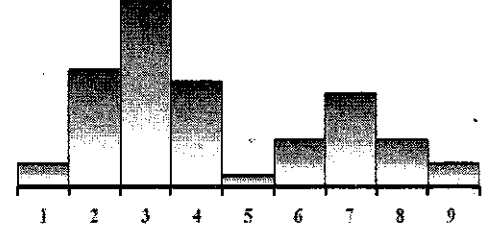
skewed left



uniform



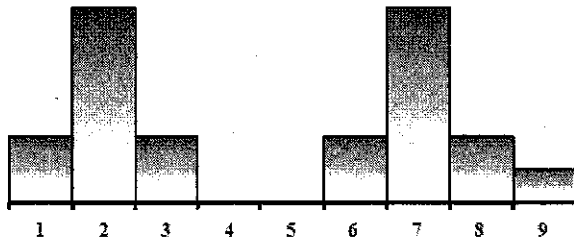
symmetric, bimodal



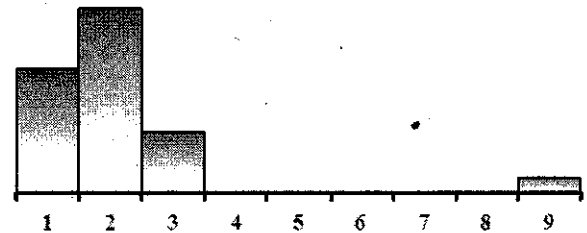
non-symmetric, bimodal

UNUSUAL FEATURES

Sometimes, statisticians refer to unusual features in a set of data. The two most common unusual features are gaps and outliers.



gap



outlier

Histogram Notes

Words to Know

Histograms

Frequency

Intervals

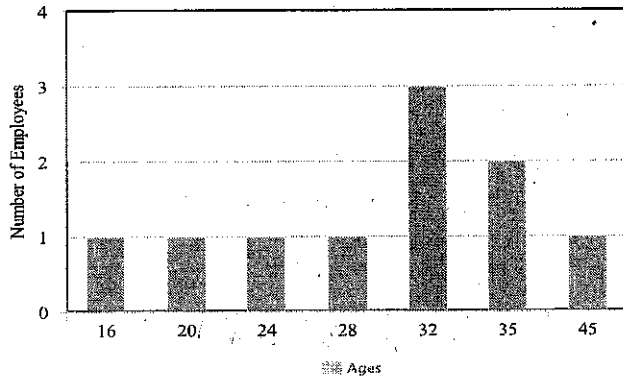
Range

The following data represents the various ages of employees at a local store.

16, 35, 20, 24, 32, 35, 45, 32, 28, 32

Bar Graph

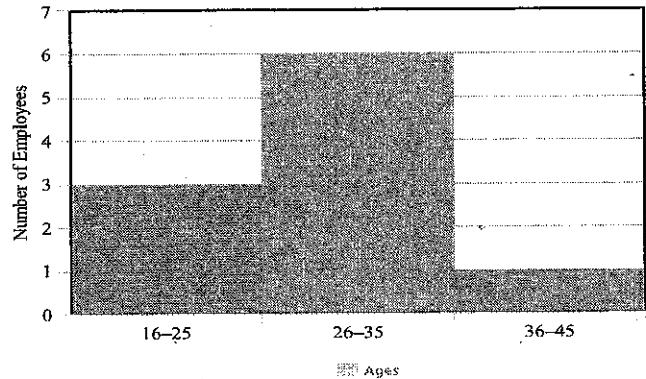
Ages of Employees at a Store



Ages	Tally	Frequency
16		1
20		1
24		1
28		1
32		3
35		2
45		1

Histogram

Ages of Employees at a Store



Intervals	Tally	Frequency
16 - 25		3
26 - 35		6
36 - 45		1

Similarities

- data shown as bars
- frequency tables
- Labeled axis

Differences

- Bars connected on histogram
- intervals (same range)

Conclusion

Histograms are very similar to bar graphs because it shows how often something occurs, known as the frequency. However, the data is grouped into intervals with the same size range. Also, the bars are connected.

Histograms are better to use than bar graphs when the data ...

large data set



Histograms

Name _____

Create a histogram for each set of data. For the first two problems, the frame is set up for you. For #3, YOU will need to determine the best ways to number the axes. Don't forget to include a title as well!

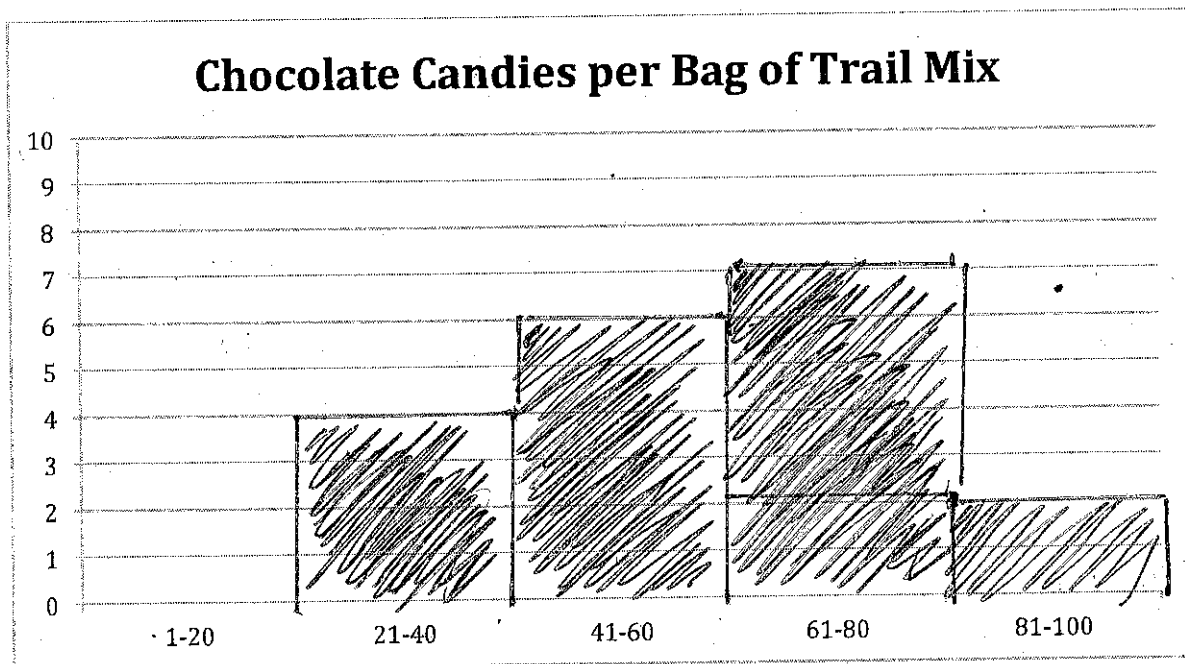
1. Chocolate candies per bag of trail mix:

50 42 119 45 68 32 67 111 61 31 75
39 62 64 49 55 51 33 117 96 64 82

Frequency table:

Interval	# of values
1-20	
21-40	4
41-60	6
61-80	7
81-100	2

organize
data
into table



119
111
117

*Describe shape and spread.

2. Test scores, out of 100 points

~~92~~ ~~84~~ ~~95~~ ~~77~~ ~~74~~ ~~80~~ ~~95~~ ~~70~~ ~~66~~
~~73~~ ~~68~~ ~~90~~ ~~78~~ ~~64~~ ~~72~~ ~~78~~ ~~76~~ ~~65~~
~~59~~ ~~71~~ ~~77~~ ~~92~~ ~~91~~ ~~89~~ ~~74~~ ~~76~~ ~~90~~

Frequency table:

Interval	# of values
30-39	
40-49	
50-59	I
60-69	IIII
70-79	IIIIIIII
80-89	III
90-99	IIII

