Objectives:
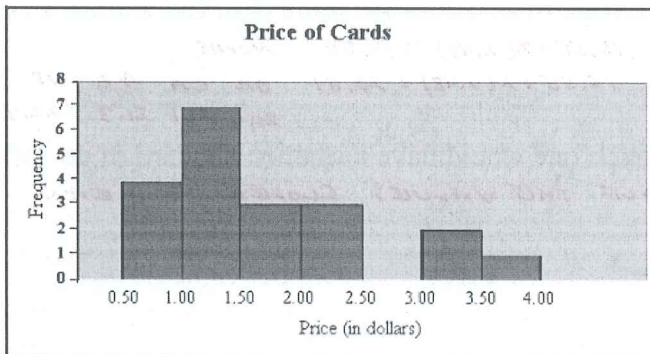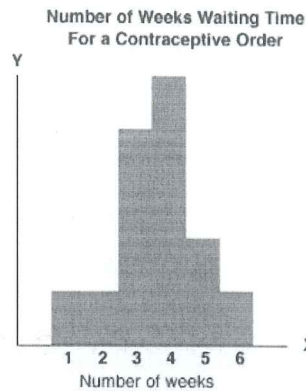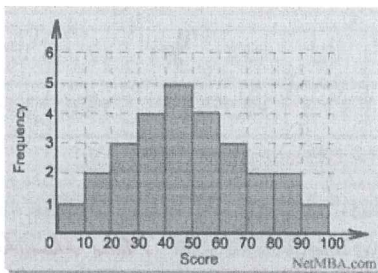1. Distinguish bar graphs from histograms.
2. Find the approximate number of items, the range, and median by studying a histogram.
3. Understand the effects of different bin widths in a histogram.
4. Create box plots and histograms in order to analyze specific characteristics of a data set.
5. Connect percentile rank and standard deviation.

Below are some graphs can you identify the odd one?

Number of Weeks Waiting Time
For a Contraceptive Order

Price of Cards

Students' Favorite Juices
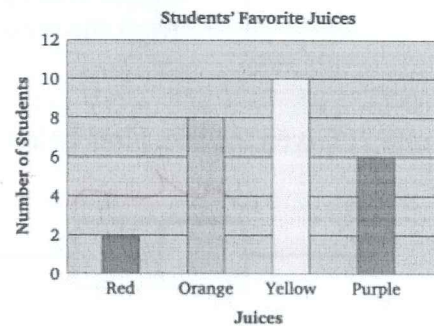
Summarize how you can tell the difference between a bar graph and a histogram in the space below.

A bar graph tells how many of each category exist, while a histogram tells how many numerical data values fall within a certain interval.

As a whole group, find the approximate number of data items, the range, and the median by studying the top left graph above.

$1 + 2 + 3 + 4 + 5 + 4 + 3 + 2 + 2 + 1 = 27$ data items

range! Anywhere from $99.\overline{9} - 0$ to $90 - 9.\overline{9}$

≈ 100 ≈ 80

median : somewhere between 40-50

Then as a small group do the same for the graph on the bottom left of the page above.

$4 + 7 + 3 + 3 + 2 + 1 = 20$ data items

range: $3.\overline{9} - .5 \approx 3.5$ to $3.5 - .\overline{9} \approx 2.5$

Median: somewhere between 1.00 - 1.5

What is the significance of bin width in a histogram? Use the words pancake and skyscraper along with some sketches to explain this below.

Smaller bin width means the bin contains a smaller range of numbers, which will show more detail, including gaps and clusters. You would also expect lower bin heights - less skyscrapers and more pancakes.

Shatevia took a random sample of 50 students who own MP3 players at her high school and asked how many songs they have stored. The two graphs were constructed from the data in the table.
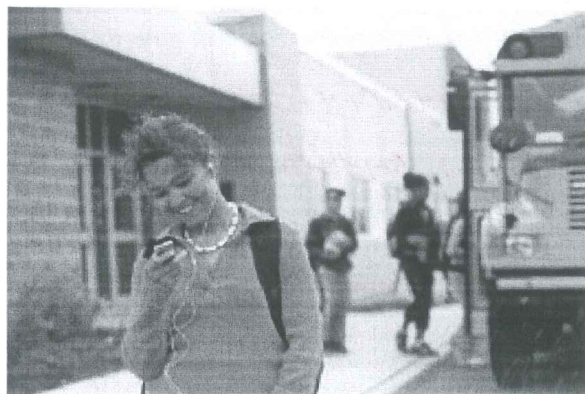
Number of Songs Stored In MP3 Player

921	866	943
796	933	900
976	841	901
905	925	863
834	895	1013
987	975	891
933	875	833
898	926	966
885	956	934
935	846	904
975	864	765
863	924	906
806	944	915
864	812	862
927	1004	925
874	974	787
896	794	

Graph A
Number of Songs in MP3 Player
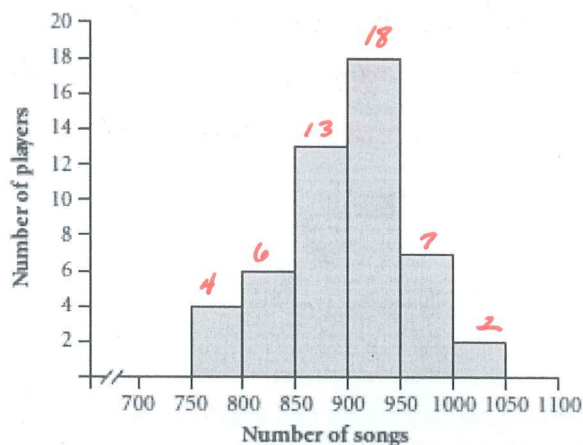
Graph B
Number of Songs in MP3 Player
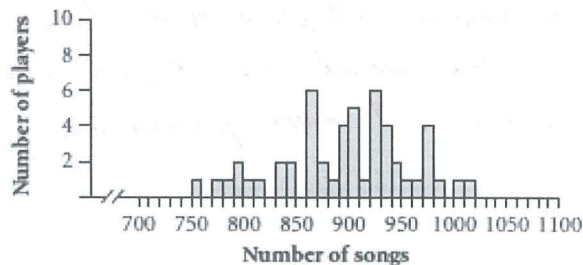
a. What is the range of the data? $1013 - 765 = 248$

b. What is the bin width of each graph? Graph A = 50   Graph B = 10

c. How can you know if the graph accounts for all 50 values? *The sum of all frequencies is 50.*

d. Why are the columns shorter in Graph B? *The bins in graph A hold the values of up to 5 bins from Graph A. Smaller bin width means shorter bins.*

e. Which graph is better at showing the overall shape of the distribution? What is that shape? *Graph A indicates that the data is skewed left (spread out more on left). This is harder to see with all the ups and downs in Graph B.*

f. Which graph is better at showing the gaps and cluster in the data? *With more bins, you can see the dots and clusters as it begins to look like a dot plot.*

g. What percentage of the players have fewer than 850 songs stored?

$^{10}/_{50} = 20\%$

Percentile Rank-

the percentage of data values that are below the given value.

In the example, 850 songs has a PR of 20 because this value is greater than 20% of values.

Percentile Example 1

Mark scored 78 on a test. The results for the whole class are shown below.

98 77 100 78 78 93 63 50 91 91 98 97 82 59

78 45 40 74 63 91 74 57 53 47 96 47 44 73

In what percentile is Mark's score?

$\dfrac{15 \text{ below}}{28 \text{ total}} = .5357$

$\boxed{53.6\%}$

Percentile Example 2

You know that 191 of the 230 students that took a test scored lower than 75. What percentile does the score of 75 represent?

$$191/230 = 83.0\%$$

Percentile Example 3

The following are the scores of 36 students on a recent algebra test. If Laura got a 74 on the test, what percentile would that score represent?

| 47	57	62	72	84	45
47	90	59	95	74	74
69	63	63	45	48	47
52	74	71	50	92	60
87	53	56	95	94	82
90	54	50	72	66	83

$$23/36 = 63.9\%$$

EXAMPLE B

The data used in this histogram have a mean of 34.05 and a standard deviation of 14.68.

a. Approximate the percentile rank of a value two standard deviations above the mean.

b. Approximately what percentage of the data values are within one standard deviation of the mean?
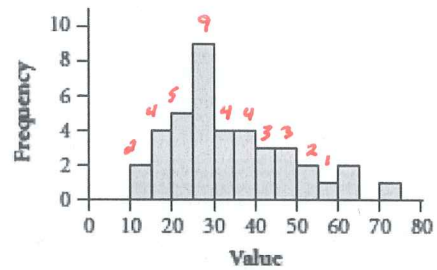
(work for *a*)

$$34.05 + 2(14.68) = 63.41$$

ALL THOSE UP TO 60: $\dfrac{2+4+5+9+4+4+3+3+2+1}{40} = \dfrac{37}{40} \approx \boxed{92.5\%}$

(work for *b*)

$$34.05 + 14.68 = 48.73$$
$$34.05 - 14.68 = 19.37$$

BINS 20 − 45: $\dfrac{5+9+4+4+3}{40}$ $\dfrac{25}{40} \approx \boxed{62.5\%}$