# Section 1.1-1.2 Notes: Describing Distributions

In this course you will be asked to **describe** or **interpret** data displays (usually *quantitative* displays).
To do this, just remember to use your **S.O.C.S.**!
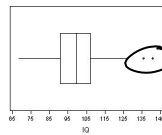
Shape Outliers Center Spread

Shape

What do I look for when trying to describe shape?

skewedness, approx normal
symmetry uniform, bimodal (peaks)

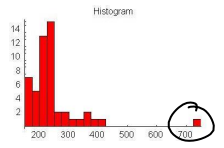*approx normal*
symmetric

tail

skew left   *uniform*   *skew Right*   *bimodal*
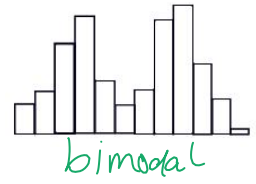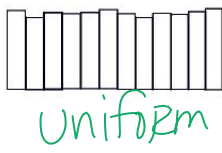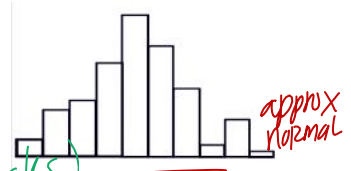
Outliers

An **outlier** is… *an extreme value that differs greatly from other points*

Stems	Leaves
1	1, 2
2	0, 2, 2, 4, 8
3	
4	
5	4

Histogram

If you are given data, you can actually calculate for outliers!

Calculating Outliers:

① Find $IQR = Q3 - Q1$
② Multiply to Find Upper & Lower Fence
LF: $Q1 - 1.5(IQR)$
UF: $Q3 + 1.5(IQR)$
any value that falls above or below those #'s

Example: Test Score Data. Are there outliers?
10, 67, 72, 75, 78, 81, 81, 82, 105

$LF = 51.5$

$UF = 99.5$

10 & 105 are outliers

1

Center

What do I look for when trying to describe center?

centric value that seems to balance all the values of graph

What numerical values are used to describe the center?

1. mean — $\bar{x}$ — average
   *not resistant to outliers
      *skew → mean is pulled in direction of skew

2. median
   middle # when in ORDER   *resistant to outliers

Spread

What do I look for when trying to describe spread?

What numerical values are used to describe the spread?

1.

2.

Description Template:
Need help describing a set of data? Check out this template to help you get started!

The distribution of _____ is _____, indicating _____. There _____
                  (variable)   (shape)   (interpret in context)   (are/are no)

outliers, as verified by the 1.5 IQR Rule.   [If there are outliers, name them and show the verification calculations.]

The typical _____ is _____, as measured by the _____. The distribution of values _____
         (variable)   (value & units)   (mean/median)   (are/are not)

consistently spread about the _____, as indicated by the _____ of _____.
                    (mean/median)   (stdev/IQR)   (value)

Mean? Median? IQR? Standard Deviation? Which one do I use???

Use these numerical descriptions…	if….

Connection to Shape and Center:
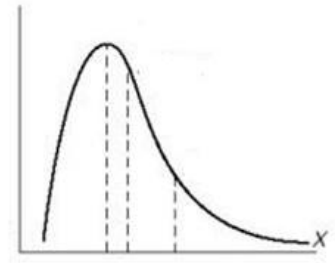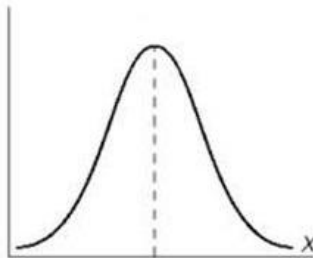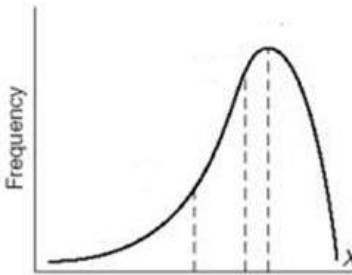
Comparing Data
You will be asked to compare the distributions of 2 more sets of data. How do you do that?

Example Time!

1. The table lists the number of home runs for each American League baseball team in 1989.

94	101	108	116	117	122	126
127	127	129	130	134	142	145

Sketch a data display for the data above and interpret the data.

2. The Chapter 1 test scores for Class 4A and Class 4B are show below.

```
            4A                4B

        9  6  4 | 5 | 1  5
     8  7  5  2 | 6 | 4  8
           4  1 | 7 | 4  6  6  7
        1  1  0 | 8 | 6  6  6  6
        2  2  1 | 9 | 4  6  7  8  9  9      5 | 1 represents 51
```

In context, compare the distributions of test scores for the two classes.

Linear Transformations

Say a class took a test and you analyzed all the test scores. After you analyzed them, finding the center, shape, spread, and outliers, the teacher decided to add 5 points to everyone's scores. How would this effect the analysis? What if the teacher added 5% to every score instead?

Analysis	Original Grade	Adding 5 Points	Adding 5%
	75		
	80		
	89		
	92		
	79		
	95		
	76		
	85		
	84		
	97		
Mean			
Median			
Standard Deviation			
Q1			
Q3			
IQR			
Range			

So what can we conclude about the analysis when we transform the data?

If this is done to the data…	These values will be affected…

More Examples!

1. Suppose that a student scores 68 on his AP Exam. In an effort to standardize grades, the grades are "curved" so that 22 points was added to each student's score, raising this student's grade to 90. Indicate how curving the grades would affect each of the following summary statistics.

   a.   Mean, Median, Range, IQR, Standard deviation

   b. How would the statistics listed in part (a) change if the students received a 10% curve?

   c. Suppose the table below shows the values of statistics listed in part a. Fill in the table for when 22 points are added to the students scores and for when 10% is added to the student scores.

	Old Value	+ 22 points	+ 10%
Mean	85		
Median	80		
Range	42		
IQR	23		
Standard Deviation	5.4		