Chapter 6: Analyzing Univariate Data and Plots

Utah Core Standards for Mathematics Correlation:

S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).
S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.
S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for effects of extreme data points (outliers).

I CAN STATEMENTS:

6.1 I can calculate mean, median, and mode. I can create and interpret box, histogram, and dot plots.
6.2 I can determine and interpret standard deviation.
6.3 I can interpret the shape of data plots.
6.4 I can interpret and compare two data sets.

Helpful Resources for this Chapter

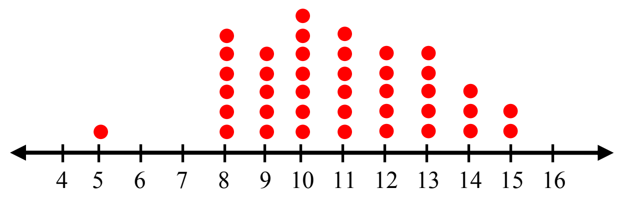
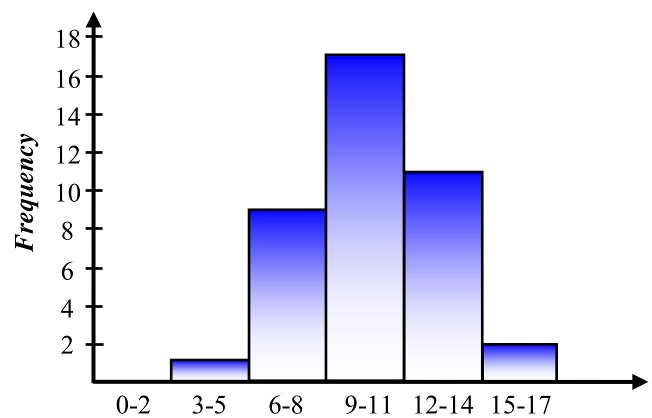
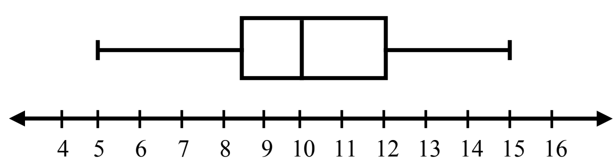
Mean-median-mode
http://www.purplemath.com/modules/meanmode.htm
http://www.regentsprep.org/regents/math/algebra/AD2/measure.htm
Create-Interpret Box, Histogram, and Dot Plots
http://www.slideshare.net/Renegarmath/box-plots-and-histograms
http://flowingdata.com/2008/02/15/how-to-read-and-use-a-box-and-whisker-plot/
video: http://www.youtube.com/watch?v=feV6j8afKz0,
video: http://www.youtube.com/watch?v=F7BZkDR3Yeg
Determine and interpret standard deviation
http://www.mathsisfun.com/data/standard-deviation.html
http://www.ltcconline.net/greenl/courses/201/descstat/mean.htm
Interpret and compare two data sets
http://openlearn.open.ac.uk/mod/oucontent/view.php?id=398296&section=1.1.3
http://www.kscience.co.uk/as/module5/ttest.htm

6.1-6.3 Interpret Data in Box, Histogram, and Dot Plots

Represent data with number line plots, including dot plots, histograms, and box plots.

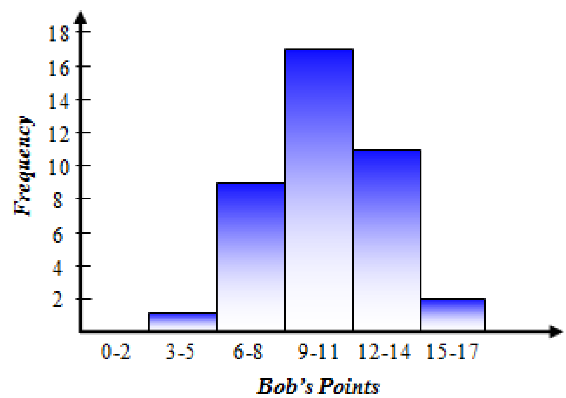
Suppose Bob scores the following points in a season of basketball games:
8, 15, 10, 10, 10, 15, 7, 8, 10, 9, 12, 11, 11, 13, 7, 8, 9, 9, 8, 10, 11, 14, 11, 10, 9, 12, 14, 14, 12, 13, 5, 13, 9, 11, 12, 13, 10, 8, 7, 8

His scores can be represented in three different ways:

dot plot: Includes all values from the range of the data and plots a point for each occurrence of an observed value on a number line.	
histogram: A histogram is a special type of bar graph. The horizontal axis represents a range of values, called an interval, instead of a single value or category—e.g., intervals of 0-2, 3-5, 6-8, 9-11, 12-14, and 15-17. The vertical axis represents the frequency of data values in the interval.	
box plot (box-and-whisker plot): A diagram that shows the five-number summary of a distribution. The summary includes the minimum, lower quartile (25th percentile), median (50th percentile), upper quartile (75th percentile), and the maximum. In a modified boxplot, the presence of outliers can also be illustrated.	

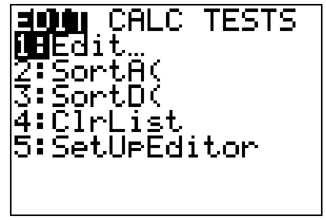
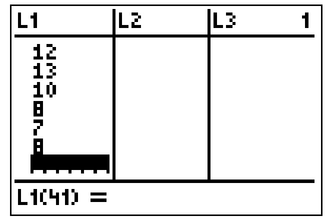
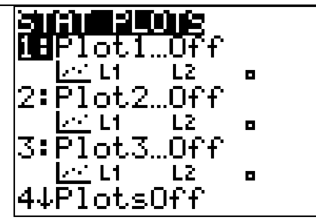
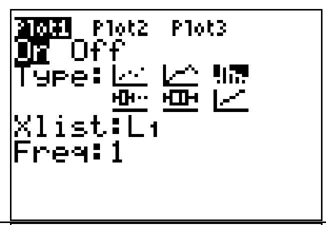
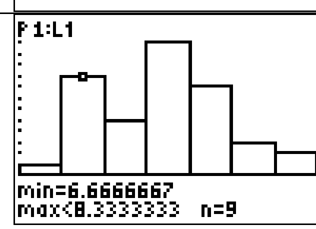
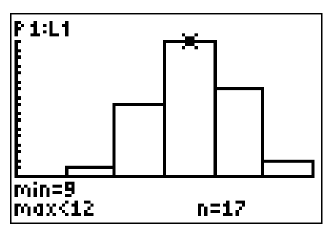
Creating a Histogram

<table>
<tr><td>Look at Bob’s scores. Determine the minimum and maximum values.</td><td>8, 15, 10, 10, 10, 15, 7, 8, 10, 9, 12, 11, 11, 13, 7, 8, 9, 9, 8, 10, 11, 14, 11, 10, 9, 12, 14, 14, 12, 13, 5, 13, 9, 11, 12, 13, 10, 8, 7, 8</td></tr>
<tr><td>Calculate the range of the data. The range is the maximum values minus the minimum value.</td><td>$15-5=10$</td></tr>
<tr><td>In our example, the length of each interval is going to be 2. Complete a frequency table with this length of interval.

You will generally be told what number of intervals to use.</td><td><table>
<tr><th>Interval</th><th>Frequency</th></tr>
<tr><td>0-2</td><td>0</td></tr>
<tr><td>3-5</td><td>1</td></tr>
<tr><td>6-8</td><td>9</td></tr>
<tr><td>9-11</td><td>17</td></tr>
<tr><td>12-14</td><td>11</td></tr>
<tr><td>15-17</td><td>2</td></tr>
</table></td></tr>
<tr><td>The last step is to graph the data from the frequency table.</td><td>
</td></tr>
</table>

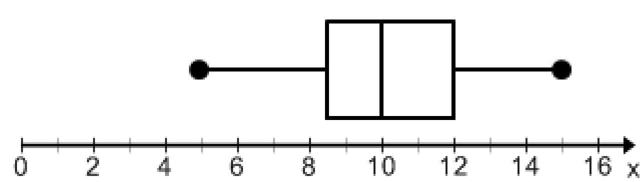
Creating a Histogram on the Calculator

Use Bob's scores: 8, 15, 10, 10, 10, 15, 7, 8, 10, 9, 12, 11, 11, 13, 7, 8, 9, 9, 8, 10, 11, 14, 11, 10, 9, 12, 14, 14, 12, 13, 5, 13, 9, 11, 12, 13, 10, 8, 7, 8

Clear your lists by pushing 2ND MEM + and selecting option L4 4. Enter Bob's scores in a list by pushing LIST STAT. This will bring up the menu shown at the right. Select Edit by either pushing L1 1 or ENTRY SOLVE ENTER. This will bring up the lists so that data can be entered. In L1 enter all of Bob's scores pushing ENTRY SOLVE ENTER after each one. Once all 40 of the scores have been entered, go back to the home screen by pushing 2ND QUIT MODE.	 
Go to the STAT PLOTS menu by pushing 2ND STAT PLOT F1 Y=. Push L1 1 or ENTRY SOLVE ENTER to select Plot1. Turn Plot1 on by pushing ENTRY SOLVE ENTER when On is highlighted. Use your arrow keys to arrow over to the histogram graphic and push ENTRY SOLVE ENTER to select it. Make sure the Xlist says L1. Go back to the home screen by pushing 2ND QUIT MODE.	 
To have the calculator automatically select the interval length, push FORMAT F3 ZOOM and arrow down to ZoomStat or push W 9 Q. This will graph the histogram with 7 intervals. By pushing CALC F4 TRACE and using your arrow keys you can see the number of data values in each interval as well as the interval. If you want to use a different interval length, you will need to change your window. To get a histogram like our example, use the following window: Xmin = 0, Xmax = 18, Xscl = 3, Ymin = -5, Ymax = 20, Yscl = 1.	 Calculator Window  Manual Window

Creating a Box-and-Whisker Plot

In order to create a box-and-whisker plot, you need the minimum, the first quartile or lower quartile, the median, the third quartile or upper quartile, and the maximum. You also need to see if there are any outliers. Outliers are more than 1.5 times the length of the interquartile range away from the first or third quartile (explanation below).

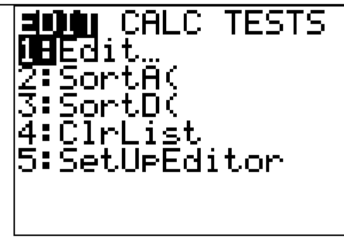
Look at Bob's scores from a season of basketball.	8, 15, 10, 10, 10, 15, 7, 8, 10, 9, 12, 11, 11, 13, 7, 8, 9, 9, 8, 10, 11, 14, 11, 10, 9, 12, 14, 14, 12, 13, 5, 13, 9, 11, 12, 13, 10, 8, 7, 8
Order the numbers from least to greatest.	5, 7, 7, 7, 8, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10, 10, 10, 11, 11, 11, 11, 11, 12, 12, 12, 12, 13, 13, 13, 13, 14, 14, 14, 15, 15
Find the median of the entire data set. There are 40 data points, so you will have to find the average of the two numbers in the middle.	$\frac{10+10}{2}=10$
Divide the list of numbers in half. Find the median of the first half to get the first quartile. Find the median of the second half to get the third quartile.	Half 1: 5, 7, 7, 7, 8, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10 Half 2: 10, 10, 11, 11, 11, 11, 11, 12, 12, 12, 12, 13, 13, 13, 13, 14, 14, 14, 15, 15 Quartile 1: $\frac{8+9}{2}=8.5$ Quartile 3: $\frac{12+12}{2}=12$
The interquartile range (IQR) is the distance between the first and third quartiles. To see if there are any outliers multiply the IQR by 1.5. See if there are any numbers less than Q1 – 5.25 or greater than Q3 + 5.25. There are no outliers.	$IQR=Q3-Q1=12-8.5=3.5$ $1.5(IQR)=1.5\cdot 3.5=5.25$ $Q1-1.5(IQR)=8.5-5.25=3.25$; the minimum is 5 $Q3+1.5(IQR)=12+5.25=17.25$; the maximum is 15
The minimum is 5, Q1 is 8.5, the median is 10, Q3 is 12 and the maximum is 15. Draw a number line and plot these values. Make a box from Q1 to Q3. Draw a vertical line in the box for the median. Connect the max and min with a whisker.	

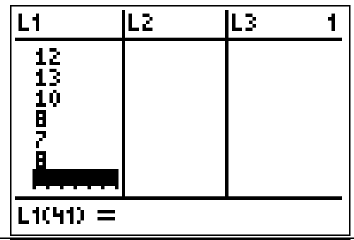
Using the Calculator to Calculate the 1-Variable Statistics

Bob scores the following points in a season:
8, 15, 10, 10, 10, 15, 7, 8, 10, 9, 12, 11, 11, 13, 7, 8, 9, 9, 8, 10, 11, 14, 11, 10, 9, 12, 14, 14, 12, 13, 5, 13, 9, 11, 12, 13, 10, 8, 7, 8

You don't have to find quartile 1, the median, or quartile 3 by hand. You can have the calculator compute the 1-variable statistics that provides a summary of all the information you may need on one set of data.

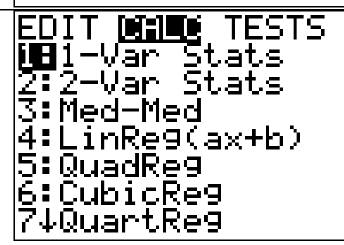
Clear your lists by pushing 2ND MEM + L4 4 T. Enter Bob's scores in a list by pushing LIST STAT. This will bring up the menu shown at the right. Select Edit by either pushing L1 1 Y or ENTRY SOLVE ENTER. This will bring up the lists so that data can be entered. In L1 enter all of Bob's scores pushing ENTRY SOLVE ENTER after each one.

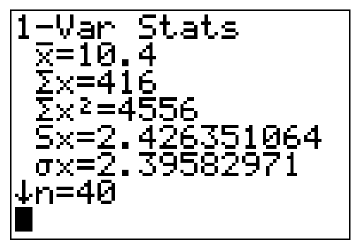


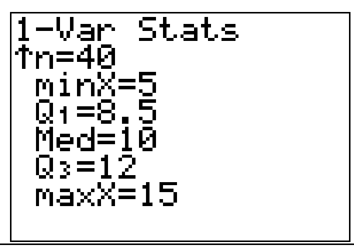


To get a summary of the data, push LIST STAT and arrow over to the CALC menu. Select the first option 1-Var Stats by pushing L1 1 Y or ENTRY SOLVE ENTER. Push ENTRY SOLVE ENTER one more time and you will get the summary. You can get more information by using your arrow keys to arrow down the screen.

$\bar{x}$ is the mean
$\sum x$ is the sum of all the data points
$\sum x^2$ is the sum of the square of all the data points
Sx is the sample standard deviation
σx is the population standard deviation
n is the number of data points
minX is the minimum
Q1 is the first quartile
Med is the median
Q3 is the third quartile
maxX is the maximum





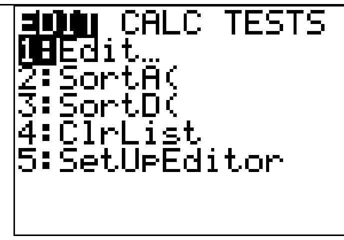


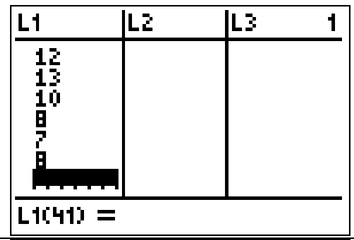
Creating a Box-and-Whisker Plot on the Calculator

Bob scores the following points in a season:
8, 15, 10, 10, 10, 15, 7, 8, 10, 9, 12, 11, 11, 13, 7, 8, 9, 9, 8, 10, 11, 14, 11, 10, 9, 12, 14, 14, 12, 13, 5, 13, 9, 11, 12, 13, 10, 8, 7, 8

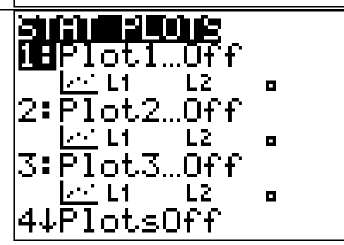
Clear your lists by pushing 2ND + 4. Enter Bob's scores in a list by pushing STAT. This will bring up the menu shown at the right. Select Edit by either pushing 1 or ENTER. This will bring up the lists so that data can be entered. In L1 enter all of Bob's scores pushing ENTER after each one.

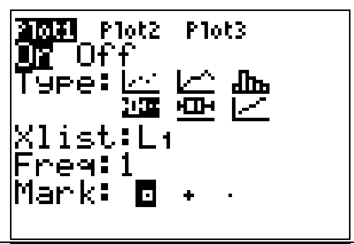
Once all 40 of the scores have been entered, go back to the home screen by pushing 2ND MODE.



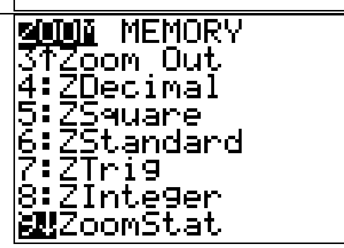


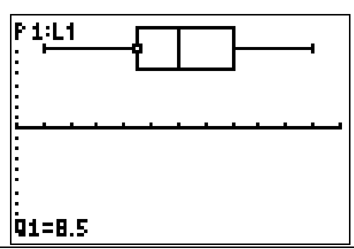
Go to the STAT PLOTS menu by pushing 2ND Y=. Push 1 or ENTER to select Plot1. Turn Plot1 on by pushing ENTER when On is highlighted. Use your arrow keys to arrow over to the first box-and-whisker graphic and push ENTER to select it. Make sure the Xlist says L1, if it doesn't push 2ND 1.





To get a nice viewing window for the box-and-whisker plot, push ZOOM and arrow down to ZoomStat or push 9. This will take you to the box-and-whisker plot. By pushing TRACE and using your arrow keys, you can see the five number summary of the box-and-whisker plot: the minimum, Q1, median, Q3, and the maximum.





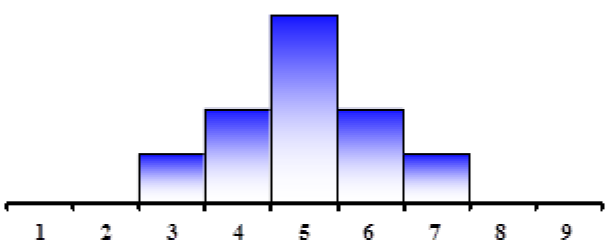
6.4 Interpret and Compare Two Data Sets

Use the shape of the data to compare two or more data sets using center (mean, median) and/or spread (range, interquartile range, standard deviation).

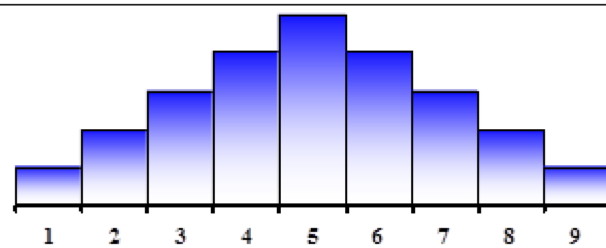
Center
Common measures of center are the median and the mean.

<table>
<tr>
<td>Median: To find the median, we arrange the data in order from least to greatest. If there is an odd number of data points, the median is the middle value. If there is an even number of data points, the median is the average of the two middle values. In the example of Bob's points scored, the median value is 10 since there are an even number of data points and the two middle values were averaged.</td>
<td>Mean: The mean of a sample or a population is computed by adding all of the data points and dividing by the number of data points. Returning to the example of Bob's points scored, the mean is (8 + 15 + 10 + 10 + 10 + 15 + 7 + 8 + 10 + 9 + 12 + 11 + 11 + 13 + 7 + 8 + 9 + 9 + 8 + 10 + 11 + 14 + 11 + 10 + 9 + 12 + 14 + 14 + 12 + 13 + 5 + 13 + 9 + 11 + 12 + 13 + 10 + 8 + 7 + 8) / 40 = 10.4 points.</td>
</tr>
<tr>
<td colspan="2">Mean vs. Median
As measures of central tendency, the mean and the median each have advantages and disadvantages.
• The median is resistant to extreme values; therefore, it is a better indicator of the typical observed value if a set of data is skewed.
• If the sample size is large and symmetric, the mean is often used as the measure of center.
To illustrate these points, consider the following example. Suppose we examine a sample of 10 households to estimate the typical family income. Nine of the households have incomes between $25,000 and $75,000, but the tenth household has an annual income of $750,000. The tenth household is an extreme value or outlier. If we choose a measure to estimate the income of a typical household, the mean will greatly over-estimate the income of a typical family (because it is sensitive to extreme values) while the median will not.</td>
</tr>
</table>

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 data points fall from the center, the greater the spread or variability of the set.



Less Spread



More Spread

Measures of Spread

Range: The range is the difference between the largest and smallest values in a set of data.

In our example of Bob's points scored, the range is 10 since we subtract our minimum value of 5 from our maximum value of 15.

**The "range" of a function is different from the "range" of data.

Interquartile range: The interquartile range (IQR) is a measure of variability, based on dividing a data set into quartiles.

Quartiles divide an ordered data set into four equal parts. The values that divide each part are called the first, second, and third quartiles; and they are denoted by Q1, Q2 (median), and Q3.

- Q1 is the "middle value" in the lower half of the ordered data.
- Q2 is the median value in the set.
- Q3 is the "middle value" in the upper half of the ordered data.

The interquartile range is equal to Q3 minus Q1. The IQR of Bob's points scored is Q3 – Q1 = 12 – 8.5 = 3.5.

*Another way to think about it: The interquartile range is the distance between the 75^{th} and 25^{th} percentile. Essentially, it is the range of the middle 50% of the data.

Standard deviation: The standard deviation is a measure of how spread out the data is. It indicates the average distance from the mean to each point of the data set. The standard deviation is a positive numerical value used to indicate how widely the individual data in a group vary.

The population standard deviation is generally not known because we rarely work with the entire population's data. The sample standard deviation is an estimate of the population's standard deviation. In the sample standard deviation, the sum of squares is divided by 1 less than the sample size to account for any error in the estimation. Computation of the standard deviation is a bit tedious and can be calculated using the following formula.

Sample standard deviation

$$s = \sqrt{\frac{\sum_{i}^{n}\left(x_i - \bar{x}\right)^2}{(n-1)}}$$

n: number of data values in the set
i: represents a single data value

Population standard deviation

$$\sigma = \sqrt{\frac{\sum_{i}^{n}\left(x_i - \bar{x}\right)^2}{n}}$$

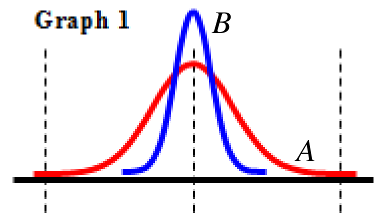
n: number of data values in the set
i: represents a single data value

The sample standard deviation for Bob's points scored is 2.4. This should be found using technology.

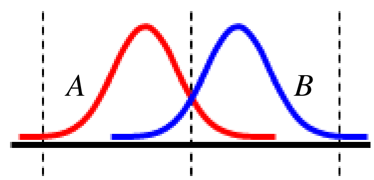
If individual data points vary a great deal from the group mean, the standard deviation is large and vice versa.

Example: In Graph 1, two sets of data are being compared. They have the same mean, but the standard deviations are different. Data A's distribution has a greater spread than data B's distribution. In Graph 2, the two distributions have about the same spread/standard deviation, but different means.

Graph 1



Graph 2



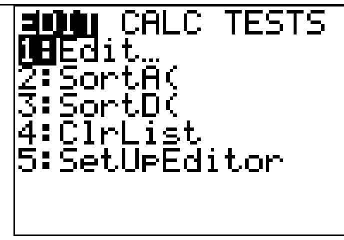
Finding Standard Deviation on the Calculator

You can calculate the standard deviation two ways on the calculator. The first way is to enter all the data in a list and calculate the 1-variable statistics (*see instructions in the box-and-whisker plot section*). This will give you both the sample standard deviation (s) and the population standard deviation (σ).

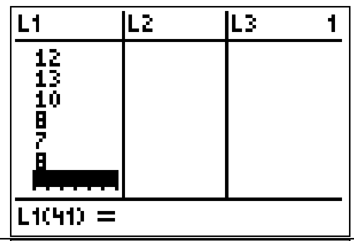
Bob scores the following points in a season:
8, 15, 10, 10, 10, 15, 7, 8, 10, 9, 12, 11, 11, 13, 7, 8, 9, 9, 8, 10, 11, 14, 11, 10, 9, 12, 14, 14, 12, 13, 5, 13, 9, 11, 12, 13, 10, 8, 7, 8

The second way is shown below:

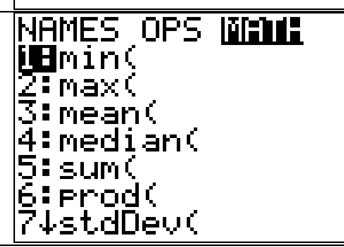
Clear your lists by pushing 2ND MEM + L4 4. Enter Bob's scores in a list by pushing LIST STAT. This will bring up the menu shown at the right. Select Edit by either pushing L1 1 or ENTRY SOLVE ENTER. This will bring up the lists so that data can be entered. In L1 enter all of Bob's scores pushing ENTRY SOLVE ENTER after each one.



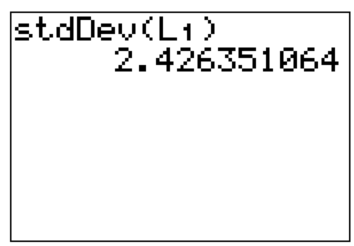
Once all 40 of the scores have been entered, go back to the home screen by pushing 2ND QUIT MODE.



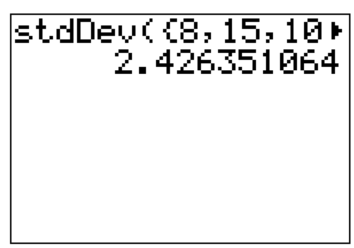
Go to the list menu by pushing 2ND LIST STAT. Use your arrow keys to arrow over to MATH. You will see that option 7 is stdDev(. You can select it by pushing u 7 or using your arrow keys to arrow down to 7 and pushing ENTRY SOLVE ENTER.



You will need to identify the list that you want to find the standard deviation of. You can identify L1 by pushing 2ND L1 1 or you can go back to the STAT menu and select L1 by pushing LIST STAT L1 1. Push ENTRY SOLVE ENTER and the sample standard deviation will be calculated.



Another way to calculate the standard deviation using this feature is to enter the data points directly instead of using a list. You need to put the data in braces and use a comma to separate each term. The braces can be found by pushing 2ND { (and 2ND }). Make sure that you begin and end the data set with the braces.

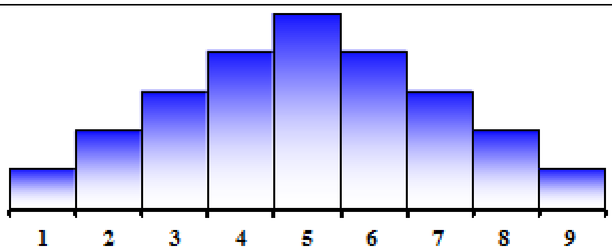


If the data has a lot of values like Bob's scores, you may want to use a list so that it is easier to modify. If there are only a few data values then you might want to just enter them using the braces.

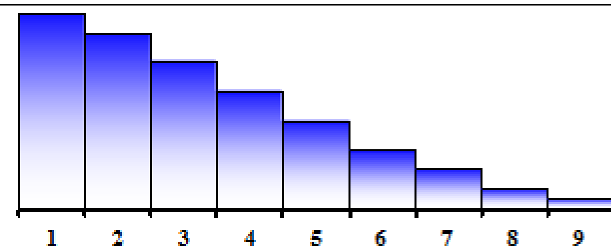
Shape

The shape of a distribution is described by symmetry, number of peaks, skewness, or uniformity.

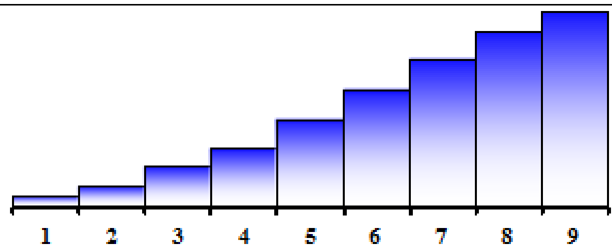
Symmetry: A symmetric distribution can be divided at the center so that each half is a mirror image of the other.	**Number of peaks:** Distributions can have few or many peaks. Distributions with one clear peak are called unimodal, and distributions with two clear peaks are called bimodal. Unimodal distributions are sometimes called bell-shaped.
Skewness: Some distributions have many more data points on one side of a graph than the other. Distributions with a tail (low levels of frequency) on the right, toward the higher values, are said to be skewed right; and distributions with a tail on the left, toward the lower values, are said to be skewed left.	**Uniformity:** When data points in a set of data are equally spread across the range of the distribution, the distribution is called uniform distribution. A uniform distribution has no clear peaks.



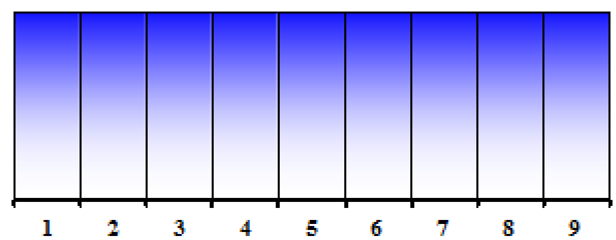
symmetric, unimodal, bell-shaped



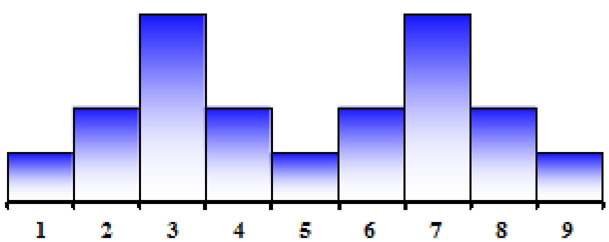
skewed right



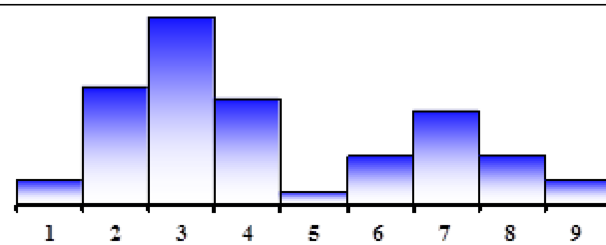
skewed left



uniform



symmetric, bimodal

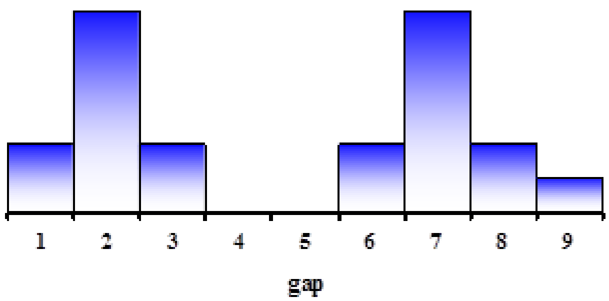


non-symmetric, bimodal

Unusual Features

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

Gaps and clusters: Gaps refer to areas of a distribution where there are no data points. The figure below has a gap between two clusters; there are no data points in the middle of the distribution.



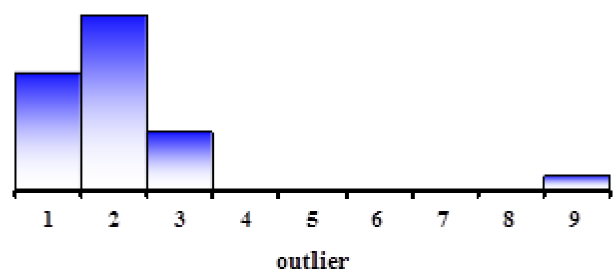
Outliers: Sometimes, distributions are characterized by extreme values that differ greatly from the other data points. These extreme values are called outliers. The figure below illustrates a distribution with an outlier. As a rule, an extreme value is considered to be an outlier if it is at least:

- 1.5 times the interquartile range below the lower quartile (Q1), or
- 1.5 times the interquartile range above the upper quartile (Q3).

OUTLIER if the values lie outside these specific ranges:

$$Q1 - 1.5 \cdot IQR$$
$$Q3 + 1.5 \cdot IQR$$

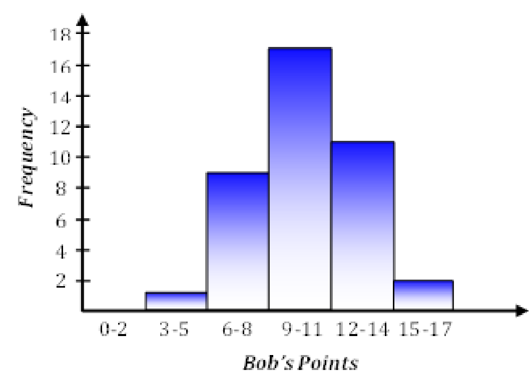
Looking back at Bob's points scored, any outliers lie outside the interval (3.25, 17.25).



Summarizing our example of Bob's points during a game:

8, 15, 10, 10, 10, 15, 7, 8, 10, 9, 12, 11, 11, 13, 7, 8, 9, 9, 8, 10, 11, 14, 11, 10, 9, 12, 14, 14, 12, 13, 5, 13, 9, 11, 12, 13, 10, 8, 7, 8

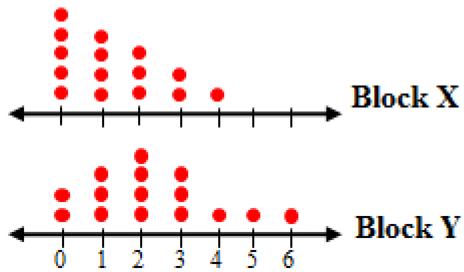
The shape is fairly symmetric. There are no gaps or outliers indicating that most of the data values are close to the center of the graph.



Comparing Distributions

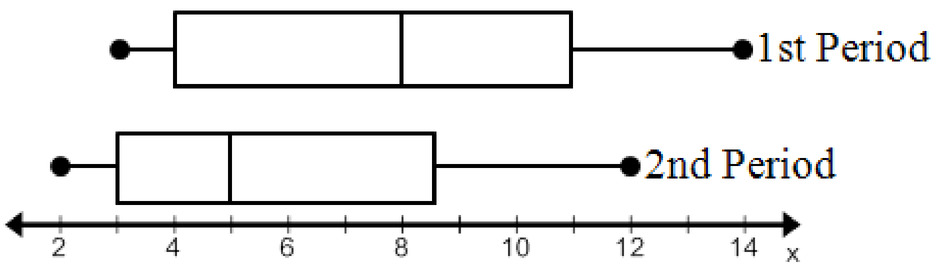
Common graphical displays such as dot plots and box plots can be effective tools for comparing data from two or more populations.

When **dot plots** are used to compare distributions, they are positioned one above the other, using the same scale of measurement.



The dot plots show pet ownership in homes on two city blocks. Pet ownership is slightly lower in Block X than in Block Y. In Block X, most of the households have zero or one pet; in Block Y, most of the households have two or more pets. Block X's pet ownership is skewed to the right while Block Y's is slightly more symmetric. In Block Y, the range of pets is from 0 to 6 pets versus 0 to 4 pets in Block X. The data distribution in Block Y is more spread out so it has more variability. There are no outliers or gaps in either set of data.

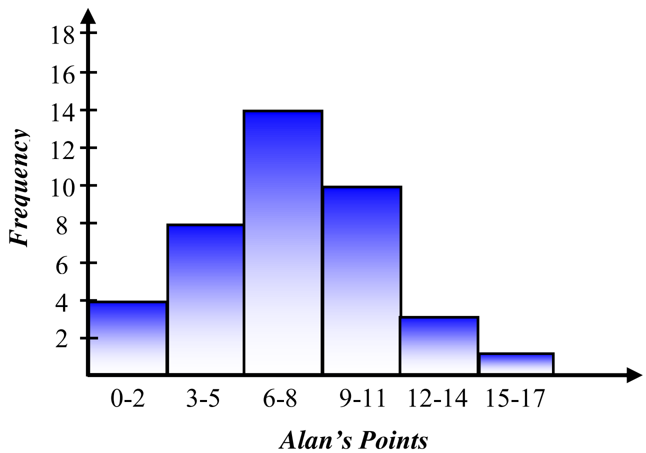
When **box plots** are used to compare distributions, they are positioned one above the other, using the same scale of measurement.

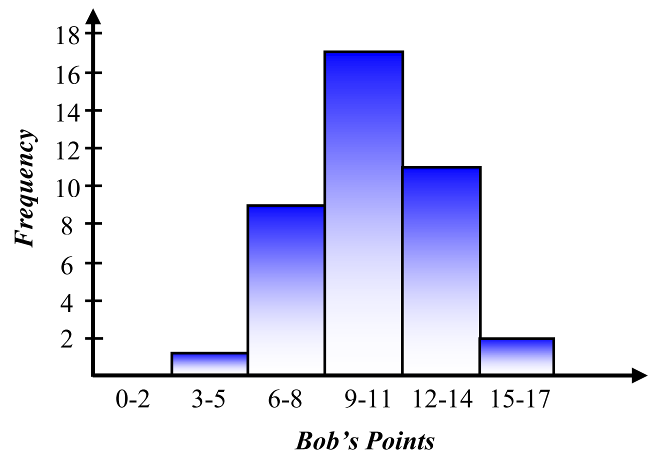


The box plots above summarize the results from a quiz given to two different class periods. The box plots show the number of questions that were answered correctly on the quiz. Neither class period had any gaps or outliers. Both distributions are a bit skewed to the right, although the skew is more prominent in the second class period. In the second class period, students got 2 to 12 questions right versus 3 to 14 questions for the first class period. While the median score is about 5 for 2nd period versus 8 for 1st period, it is difficult to say that one period did better than the other since the medians of both periods lie within the interquartile range of the other.

Suppose Bob's friend Alan had these scores: 1, 3, 0, 2, 4, 5, 7, 7, 8, 10, 4, 4, 3, 2, 5, 6, 6, 6, 8, 8, 10, 11, 11, 10, 12, 12, 5, 6, 8, 9, 10, 15, 10, 12, 11, 11, 6, 7, 7, 8

Alan's five-number summary statistics are as follows: the minimum is 0, quartile 1 is 5, the median is 7.25, quartile 3 is 10, and the maximum is 15. The mean value is 7; the range of the values is 15, the IQR is 5, and the standard deviation is about 3.4. Any outliers must be outside of the interval (-2.5, 17.5). Comparing these two sets, Alan has a larger spread of data with an IQR of 5 points versus Bob's IQR of 3.5 points. In addition, Alan's overall scores are generally lower than Bob's, shown by Alan's median 7.5 points which is less that Bob's median score of 10 points. The shape of Alan's graph is not as symmetric as Bob's, suggesting that Alan's scores might be slightly skewed right since there is a bit of a tail in that direction whereas Bob's scores appear to have a symmetric distribution.





6.0—Task: Sewing Shirts Award

Pretend that you manage the assembly line of workers sewing men's shirts in a clothing factory. Each month you give a bonus for productivity to one employee. This month, you are tracking employees who sew the sleeves. Below are the numbers of completed shirts (with sleeves) for the two most productive employees for 12 days. Use mathematics to decide which of these two top performers should get the award. You will need to explain your choice to both employees.

Day	***Adriana***	***David***
1	32	49
2	75	45
3	38	51
4	42	49
5	47	63
6	68	56
7	51	51
8	51	48
9	58	52
10	31	42
11	51	51
12	65	52

6.1—Calculate Mean, Median, and Mode. Create and Interpret Box, Histogram, and Dot Plots

Use this blank page to compile the most important things you want to remember for cycle 6.1:

6.1a (refine)—What is Average?

1. Some people use "average" interchangeably for both mean and median. Consider this statement:

   "Just think of how stupid the average person is, and then realize half of them are even stupider." George Carlin

   What type of "average" is George Carlin referring to, mean or median? Is it possible to have <u>more</u> than half of a population above this kind of average?

2. What is the difference between mean and median?

3. Give an example of data when the mean and median might have the same value.
   Give an example when the mean and the median do NOT have the same value.

4. Can the following statement be true? Why or why not?
   "Welcome to Lake Wobegon, where all the women are strong, all the men are good-looking, and all the children are above average." Garrison Keillor

5. Is it possible to have more than half of the data values above (or below) the mean? Explain why or why not.

6. So…what is average? Explain. Then explain the humor in the following statement.
   "Now there are more overweight people in America than average-weight people. So, overweight people are now average. I guess this means you've met your New Year's resolution." Jay Leno

7. Find the mean, median, and mode for each set of data.
   a. 5, 11, 16, 8, 4, 7, 15, 6, 11, 7
   b. 24, 25, 20, 21, 11, 14, 12, 12, 13

8. Kara had 85, 83, 92, 88, and 69 on her first five math tests. She knows that she needs an average of 85 to get a B. What score must she get on her last test to get a B?

9. Miss Johnson teaches 4 sections of Secondary Math I. Her overall mean on the last test was 87. If three of the class means were 92, 88, and 80, what was the approximate mean for the fourth class?

For questions 10 and 11, calculate the mean, median, and mode. Then tell which best describes the center of the data.

10. The low temperatures the second week of February in Salt Lake were 33°, 32°, 34°, 33°, 33°, 33°, 34°.

11. A given data set: 28, 580, 100, 228, 31, 192, 13, 130, 127, 301, 40, 187, 112

6.1b (apply)—Test Score Comparison

Use the following student test score data. Interpret, represent, and analyze the data according to the instructions.

<u>Class A Test Scores:</u> 51, 45, 45, 45, 33, 51, 48, 36, 48, 51, 27, 51, 36, 48, 51, 39, 51, 39, 30, 51, 39, 51, 42, 48, 33, 51, 48, 42, 45, 51, 21, 39, 51

<u>Class B Test Scores:</u> 48, 51, 48, 24, 48, 51, 48, 48, 51, 18, 48, 51, 48, 45, 21, 30, 36, 48, 45, 51, 36, 39, 30, 45, 33, 45, 27, 39

1. Find the three measures of central tendency and the lower and upper quartiles for the data.

	CLASS A	CLASS B
Mean		
Mode		
Lower Quartile		
Median		
Upper Quartile		

2. Create a histogram with 8 intervals beginning with the interval 15–20.

3. Create a dot/line plot.

4. Create a box-and-whisker plot.

5. <u>Analyze</u> – Which class did better overall? How can you tell?

6. <u>Analyze</u> – Which measurement <u>best</u> helps you to evaluate which class did better?

7. <u>Compare</u> – Which class has a higher average score?

8. <u>Compare</u>—Write a comparison about student performance in the 2 classes. Use any of the information from the table above that will help you compare.

9. <u>Compare</u> – What effects did outliers have on the data? If outliers were removed, how would it change the overall averages?

6.2—Determine and Interpret Standard Deviation

Use this blank page to compile the most important things you want to remember for cycle 6.2:

6.2a (refine)—Standard Deviation: Test Scores for Joe and Sam

A measure of distribution is a measure of how spread out data is, or how the data is distributed from its smallest values to its largest values. Suppose, for instance, that Joe has test scores of 60, 68, 69, 78, 90, 95, and 100. Sam scores 78, 78, 79, 79, 82, 82, and 82.

1. Calculate Joe's mean test score. Then calculate Sam's mean test score. What do you notice about Joe's scores compared to Sam's?

Measuring the mean will not tell you much about the characteristics of the test takers' performance. A measure of distribution, or spread, will help you see that Sam consistently scores near 80, while Joe's scores are spread out, or distributed, over a much larger range.

2. One way of examining the distribution of test scores is to find the standard deviation. Follow the steps below to find the standard deviation of Sam's test scores (Joe's example is given).

	Joe	**Sam**
1.Calculate the mean ($\bar{x}$) of the data.	*Mean:* $\bar{x}=\frac{60+68+69+78+90+95+100}{7}$, $\bar{x}=80$	
2.Find the deviation, or distance from the mean, for each piece of data.	*Deviation*: $60-80=-20$ $68-80=-12$ $69-80=-11$ $78-80=-2$ $90-80=10$ $95-80=15$ $100-80=20$	
3. Find the *mean squared deviation*, or *variance* (s^2).	*Variance for Joe's scores:* $(-20)^2=400$, $(-12)^2=144$, $(-11)^2=121$, $(-2)^2=4$, $(10)^2=100$, $(15)^2=225$, $(20)^2=400$ $s^2=\frac{400+144+121+4+100+225+400}{7}$ $s^2=\frac{1394}{7}=199.1429$	
4. Find the *standard deviation*, (σ) as the square root of the variance. Symbolically this is $\sigma=\sqrt{s^2}$.	*Standard Deviation:* $\sigma=\sqrt{s^2}=\sqrt{199.1429}\approx 14.1118$	

3. Why do you think the standard deviation of Joe's test scores is higher than the standard deviation of Sam's test scores?

4. What does the standard deviation tell you about a data set?

6.2b (apply)—Standard Deviation: Pull-Up and Car Data

(The applet for mean and standard deviation is found at:
http://www.math.usu.edu/~schneit/CTIS/SD/index.html)

Homework: For this homework assignment, you can use the applet listed above to check your calculation of the mean and standard deviation.

1. **Pull-Up Data:**
   a. A gym teacher at a middle school collected this data about the number of pull-ups by seventh graders in gym class 2, 3, 4, 3, 2, 5, 5, 6, 6, 6, 9, 4, 10, 3, 2, 1, 9
      1. Create a histogram with 4 intervals starting with the interval 0–2 to display the data.
      2. Find the mean and standard deviation. Show the steps. Then, if desired check those calculations using the standard deviation applet above.

   b. Six of the school football players were gone the day that the pull-up test was given. Their results are 7, 7, 8, 10, 10, 9.
      1. Add these points to your histogram.
      2. How did these points change the mean? How did these new data points change the standard deviation?

   c. How does finding standard deviation help you understand the data?

2. **Car Data:**
   a. The following data was observed by a city planner. The city planner recorded the number of cars that went through an intersection during a given hour 4, 5, 3, 6, 7, 8, 5, 6, 7, 2, 3, 6, 4, 4.
      1. Create a histogram with 4 intervals starting with the interval 2–3 to display the data.
      2. Find the mean and standard deviation. Show the steps. Then, if desired check those calculations using standard deviation applet above.

   b. The city planner decided that he didn't have enough data from his first set of data points so he went back to the intersection and observed a few more data points. The results of this data are 5, 9, 7, 3, 2.
      1. Add these data points to your existing graph.
      2. What are the mean and standard deviation now? How did they change?

   c. **Analysis:** What is the relationship between the spread of the data and the standard deviation?

6.3—Interpret the Shape of Data Plots

Use this blank page to compile the most important things you want to remember for cycle 6.3:

6.3a (build)—Symmetric and Skewed Distributions

For Problems 1–2

a. Create a histogram of the data in the table.

b. Find the Mean and the Standard Deviation.

c. Identify the data as symmetric, left skewed, right skewed, or other.

d. Write one sentence describing the meaning of the distribution.

1. Candies in a Bag

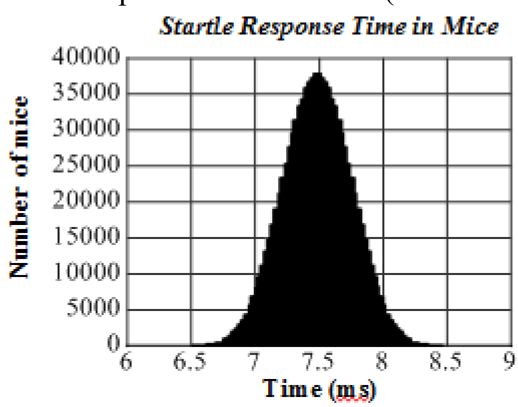
Number of Candies	*Frequency*
21	1
22	2
23	4
24	10
25	20
26	10
27	4
28	2
29	1

2. Late math assignments

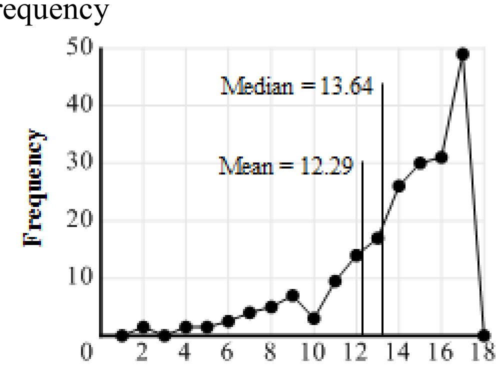
Number of Late Assignments	*Frequency*
2	10
3	12
4	10
5	30
6	28
7	32
8	30
9	25

For problems 3–6, identify the distribution as symmetry, left skewed, right skewed, or other.

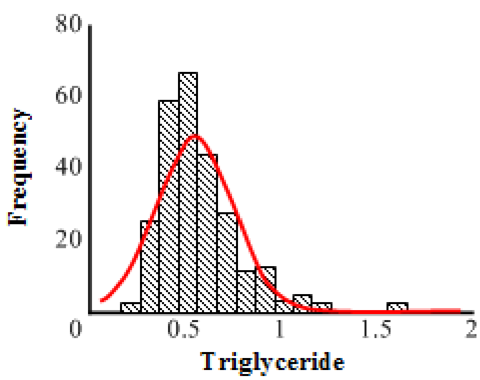
3. Startle Response Time in Mice (#mice/time)



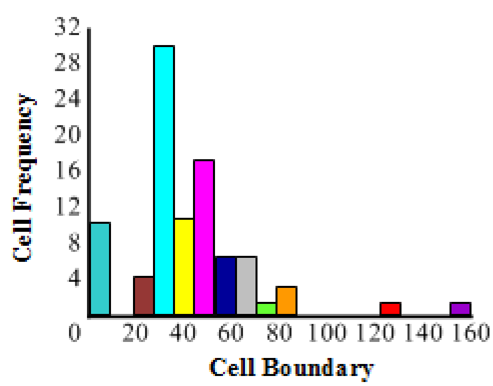
4. Frequency



5. Frequency/Triglycerides



6. Cell Frequency/Cell Boundary



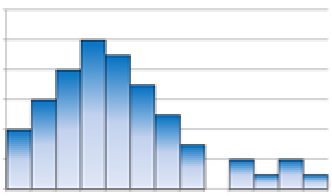
6.3b (refine)—Interpreting Statistics

For each histogram/dot plot:

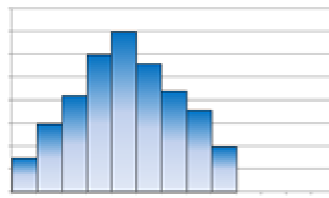
1. Describe the distribution of the data. (Shape: Symmetrical, skewed left, skewed right, uniform, or other)

2. Looking at the most common measures of center, determine whether the mean of the histogram is greater than, less than, or about the same as the median.

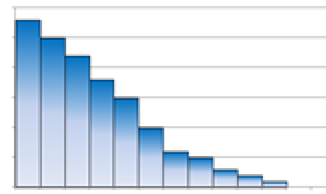
A.



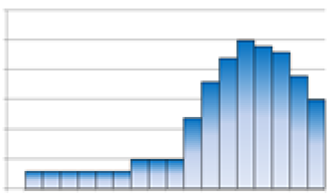
B.



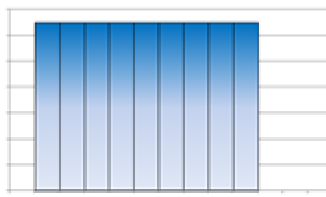
C.



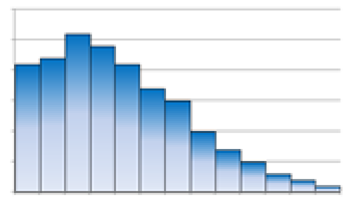
D.



E.



F.



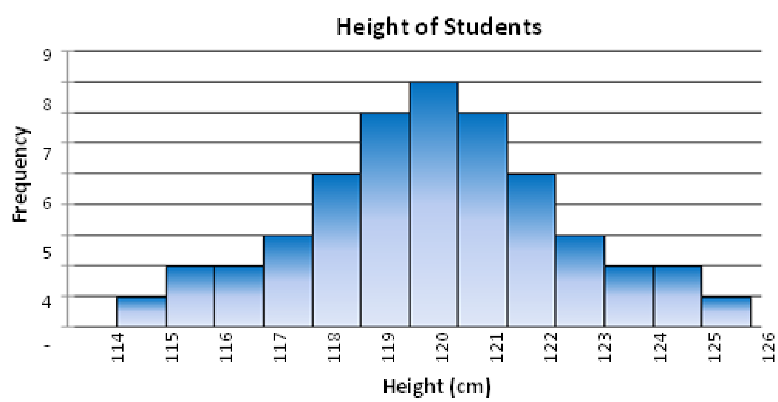
G.

Use the graph to the right to answer questions 3—5.

3. Draw a curve that follows the data distribution.

4. This graph is an example of a ______________ distribution.

5. What can be said about the mean and median of this data set?



Use the graph to the right to answer questions 6–13.

6. What kind of distribution are both sets of bear data?

7. Which graph has a smaller standard deviation?

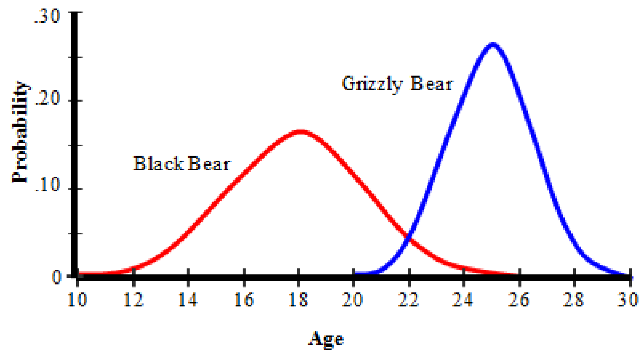
8. Which graph has a larger range of data?

9. Which bear tends to live longer?

10. What is the average age of a black bear?

11. What is the mean and median of the grizzly bear?

12. Are the two spreads equal? Explain.



6.4—Interpret and CompareTwo Data Sets

Use this blank page to compile the most important things you want to remember for cycle 6.4:

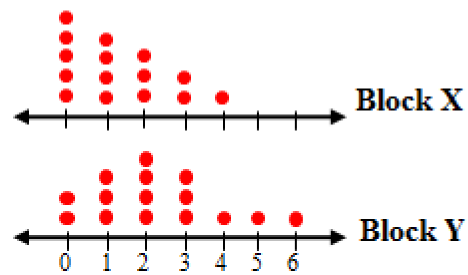
6.4b (refine)—Interpreting Plots Practice

The dot plots below show pet ownership in homes on two city blocks.

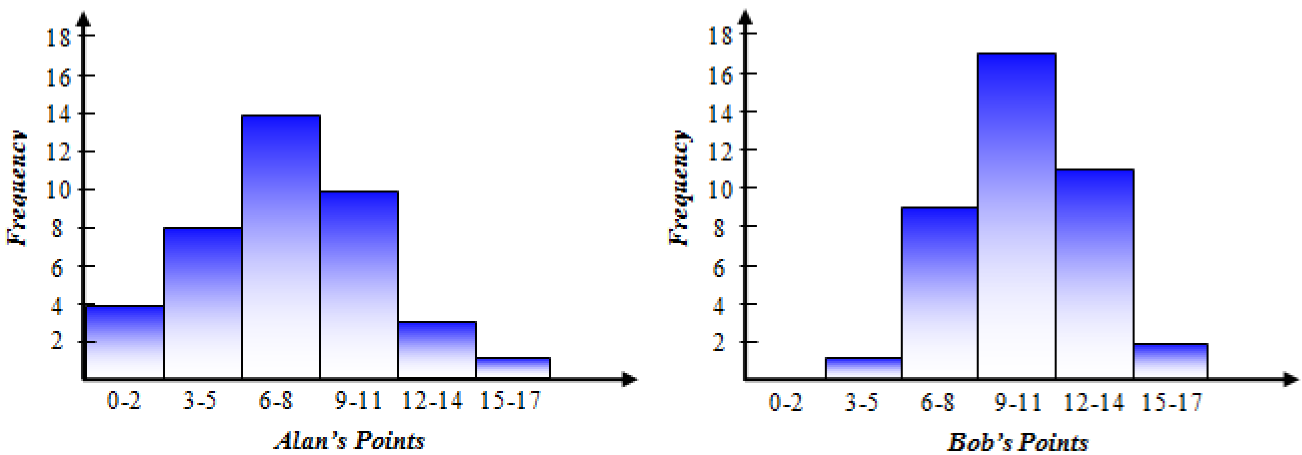
1. Compare the two dot plots at right. Use the following questions to help compare.

Explain

a. Which block has lower pet ownership?
b. Which block is skewed right?
c. Which block appears more symmetric?
d. What is the range for each block?
e. Which block has more variability?
f. Do either of the blocks appear to have any outliers?



2. The following Histograms show points scored for Alan and Bob



a. Who has a larger spread of data? Justify your answer.
b. Which person data is more symmetric? Justify your answer.
c. Which person's data is slightly skewed right? Justify your answer.

6.4c (refine)—Compare Box-plots: NBA Salaries

The winners of the NBA Championship in 2008 were the Boston Celtics. Given are the Boston Celtics and Utah Jazz salaries for the 2008/09 year. (http://hoopshype.com/)

	Boston Celtics	***2008/09***
1	Kevin Garnett	$24,750,000
2	Paul Pierce	$18,077,903
3	Ray Allen	$17,388,430
4	Kendrick Perkins	$4,078,880
5	Brian Scalabrine	$3,206,897
6	Eddie House	$2,650,000
7	Tony Allen	$2,500,000
8	Rajon Rondo	$1,646,784
9	Stephon Marbury	$1,262,275
10	JR Giddens	$957,120
11	Leon Powe	$797,581
12	Gabe Pruitt	$711,517
13	Glen Davis	$711,517
14	Mikki Moore	$378,683
15	Bill Walker	$542,114
	TOTAL:	$79,659,701

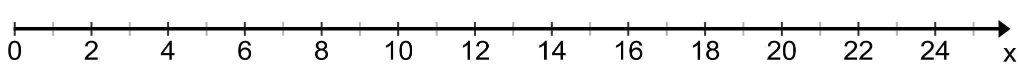
	Utah Jazz	***2008/09***
1	Andrei Kirilenko	$15,080,312
2	Carlos Boozer	$11,593,817
3	Mehmet Okur	$8,500,000
4	Matt Harpring	$6,000,000
5	Deron Williams	$5,069,448
6	Kyle Korver	$4,773,927
7	CJ Miles	$3,700,000
8	Jarron Collins	$2,074,302
9	Brevin Knight	$2,000,000
10	Ronnie Brewer	$1,834,680
11	Ronnie Price	$1,188,000
12	Kosta Koufos	$1,129,320
13	Morris Almond	$1,081,440
14	Kyrylo Fesenko	$810,000
15	Paul Millsap	$797,581
	TOTAL:	$65,632,827

1. Find the five number summary for this data: Min =__, Q1 =__, Median =__, Q3 =__, Max =__.

2. Use the number line below. Above the number line, create a box plot for the Celtics. *The number line is in millions of dollars. Example: Eddie House's salary of $2,650,000 would be 2.65 on the number line.*

   Find the mean salary for a Celtic player. Label this spot on your number line.

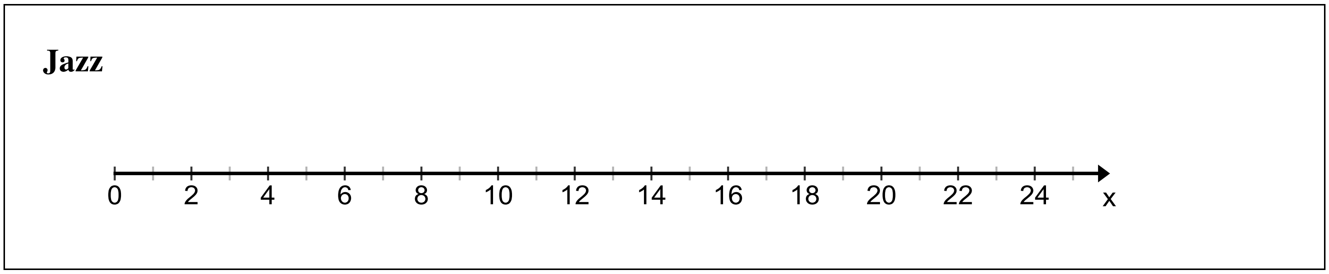
Celtics



3. Which is a more appropriate measure of the center of the salary data, the median or the mean? Explain your choice.

4. The five number summary for Utah Jazz salary data is:
   Minimum = 798 Q1 = 1129 Median = 2074 Q3 = 6000 Maximum = 15080

Create a box plot for the Jazz on the number line.



5. The mean salary for the Jazz is: ____________. Label this spot on your number line.

6. Compare the measures of center for the Celtics and the Jazz.

6.4d (apply)—Utah Jazz Stats

2012-2013 Jazz Player Weight and Heights

Player Name	Weight (pounds)	Height (inches)
Raja Bell	210	77
Alec Burks	202	78
DeMarre Carroll	212	80
Jeremy Evans	194	81
Derrick Favors	248	82
Randy Foye	213	76
Gordan Hayward	210	80
Al Jefferson	289	82
Enes Kanter	267	83
Paul Millsap	253	80
Kevin Murphy	185	78
Jamaal Tinsley	197	75
Earl Watson	199	73
Mo Williams	195	73
Marvin Williams	245	81

1. Create a box plot on the graphing calculator for Jazz player weights. Record the five number summary for the weights.

2. Create a box plot on the graphing calculator for the Jazz player heights. Record the five number summary for the heights.

3. Compare the plots for Jazz player heights and weights. What are the similarities and differences?

4. Pretend that Yao Ming (7'6" was recruited for the Jazz). Predict how adding Yao Ming will change the box plot for Jazz Heights.

5. Create a new box plot for Jazz player heights (Yao Ming included). How did the plot change? How did the five statistics in the five number summary change. Explain the changes.