11-0 NOTES: Creating Statistical Plots

	Definition	How do we find it?
Mean	Finding the average	Add up all the numbers and divide by how many numbers are listed
Median	The middle number	List the numbers in order and find the one that occurs in the middle of the list. If two numbers occur in the middle, find the average of the two numbers.
Mode	The number that occurs the most	Look for the number that occurs more than once. (It is possible for there to be more than one mode.)
First Quartile/ Lower Quartile (Q_1)	The value, which under 25% of the points are found	After you find the median, look at the lower half of the data and find the median for the lower half of the numbers listed.
Third Quartile/ Upper Quartile (Q_3)	The value, which over 25% of the points are found	After you find the median, look at the upper half of the data and find the median of the upper half of the numbers listed.
Range	The difference between the lowest and highest number	Highest number (maximum) minus lowest number (minimum)
Interquartile Range (IQR)	The difference between the lower quartile and the upper quartile.	Upper (third) quartile minus lower (first) quartile
Outlier	A value that is abnormally far away from the rest of the values.	If the number is less than $Q_1 - 1.5(IQR)$ OR if the number is greater than $Q_3 + 1.5(IQR)$

Example 1 The April high temperatures (in degrees Fahrenheit) for five consecutive years in Boston are listed below. Find the mean, median, range and interquartile range for this data set.

$$77 \qquad 86 \qquad 84 \qquad 93 \qquad 90$$

Mean:	Range:
Median:	Interquartile Range:

→**Outlier** – a statistically insignificant value because it is either much lower than or much higher than the rest of the data. *Note:* Outliers are usually denoted on plots as asterisks (*).

A number is considered an outlier if: The number $x < Q_1 - 1.5(IQR)$

Or

The number $x > Q_3 + 1.5(IQR)$

***Are there any outliers in Example 1?

Mean measures the center of the data. It is represented by $\bar{x}$, pronounced "x bar."
Standard Deviation measures the spread of the data. It is represented by σ.

<u>**Steps for finding Standard Deviation**</u>
Step 1: Calculate the mean, $\bar{x}$.
Step 2: Calculate each data value's deviation from the mean by finding $x - \bar{x}$ for each data value x.
Step 3: Find $(x - \bar{x})^2$, the square of each deviation.
Step 4: Find the sum of all $(x - \bar{x})^2$.
Step 5: Divide the sum of $(x - \bar{x})^2$ by the number of numbers, the mean of the squared deviations.
Step 6: Take the square root of the mean of the squared deviations.
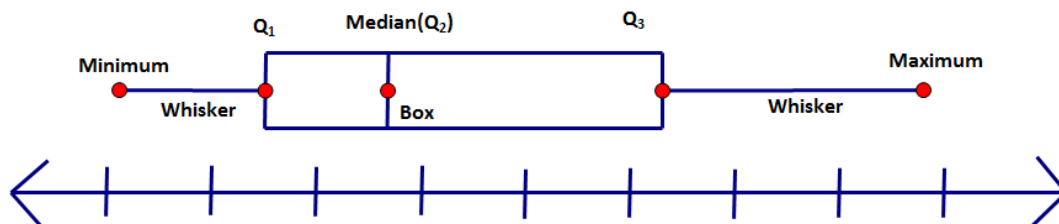
Example 5 Calculating the Standard Deviation
Calculate the standard deviation for the data from Example 1. Use the table to assist you.

Calculate the mean: $\bar{x} =$ _____

Data value, x	Deviation from mean, $x - \bar{x}$	Squared deviation, $(x - \bar{x})^2$
77		
84		
86		
90		
93		
Sum of squared deviation, $\sum(x - \bar{x})^2$		
Mean of squared deviation, $\dfrac{\sum(x-\bar{x})^2}{n}$		
Square root of mean, $\sigma =$ $\sqrt{\dfrac{\sum(x-\bar{x})^2}{n}}$		

DATA PLOTS

A **box-and-whisker plot** displays a five-number summary of data sets. Write in the percent of the data that is represented by each part of the box plot.

Create a box-and-whiskers plot
Step 1: Rank the numbers in order.
Step 2: Find the median (also called the second quartile).
Step 3: Find the first quartile (the median of the lower half).
Step 4: Find the third quartile (the median of the upper half).
Step 5: Identify the minimum and maximum (smallest and largest numbers).
Step 6: Create a number line to represent the data.
Step 7: Draw dots above the number line to represent the minimum, maximum and quartiles.
Step 8: Draw a short vertical line through the quartiles, create the box and whiskers and label the graph.
Step 9: Label the five critical points (minimum, maximum and the three quartiles).
Note: do not include outliers as part of the 5 data values. Mark them with asterisks (*) instead.

Example 2 Create a box-and-whiskers plot to represent the following monthly sales, in millions of dollars, of audio devices.   15 4 9 16 10 16 8   14 25 34

```
|——————|——————|——————|——————|——————|——————|——————|
0       5      10     15     20     25     30     35
```

A **Stem and Leaf Plot** is a way to group data.

Creating a Stem and Leaf Plot
Step 1: Find the smallest and largest values in the data.
Step 2: Draw a vertical line. To the left of the line, write each consecutive digit from the smallest stem value to the largest stem value.
Step 3: Insert the leaves.
Step 4: For each stem, reorder the leaves from smallest to largest.
Step 5: Include a key.

Example 3 Construct a stem and leaf plot for the following and answer the questions.
In September, each of the 24 students in a math class reported the number of days he or she work that summer: 25, 17, 15, 28, 24, 13, 16, 28, 19, 25, 24, 36, 33, 18, 24, 38, 28, 25, 27, 14, 37, 28, 35, 43

A **Dot Plot** is a representation of the data that shows the spread and outliers.

Creating a dot plot
Step 1: Create a number line to represent the spread of the data.
Step 2: For every piece of data, place an X above that value on the number line.
Step 3: Label the graph

Example 4 Create a dot plot to display the data
Twelve employees at a small company make the following annual salaries (in thousands of dollars):
25, 30, 35, 35, 35, 40, 40, 40, 45, 45, 50, 60, 150

0   50   100   150

After completing the dot plot fill out the table below.

Mean	Median	Range	IQR	Standard Deviation

An **outlier** is a value in a data set that is relatively much greater or much less than the other values in the data.
Determining Whether a Data Value is an Outlier:
A data value x is an outlier if $x < Q_1 - 1.5$ (IQR) or if $x > Q_3 + 1.5$(IQR)

From example 4, show that 150 is an outlier.

A **Histogram** displays the groups of data values into adjoining intervals of equal width and uses the heights of bars to indicate the number of data values (frequency) that occur in each interval.

Creating a Histogram
Step 1: Find the minimum and maximum values and create equal intervals to include these points.
Step 2: Create a frequency table and fill it in using the data.
Step 3: Draw a graph in quadrant 1 and label axes and graph.
Step 4: Create the bars of the histogram. Remember the bars must touch.

Example 7 Create a histogram to represent the following information.
The finishing times, in second, for a race are shown below. Create a histogram with 5 intervals that represents the data.

95 105 83 81 93 98 102 99 82 89 90 82 89

Times	Frequency