2024-2025 AP Stats Summer Work

Ms. Braillard
mbraillard@mvrhs.org

The purpose of this summer work is to review some basic topics in previous math classes, to introduce you to some statistical capabilities of your graphing calculator and familiarize you with the statistical language we will be using throughout the year. Statistics is unlike any other math class you have taken in your past, it has a language component that not all students are familiar with. This is a rigorous course and requires dedicated time outside of the classroom in order to fully understand the concepts presented.

TI 84 Calculator: It is required that you have a TI-84 calculator for this course. We use the calculator everyday and you must have your own calculator to be comfortable working with it.

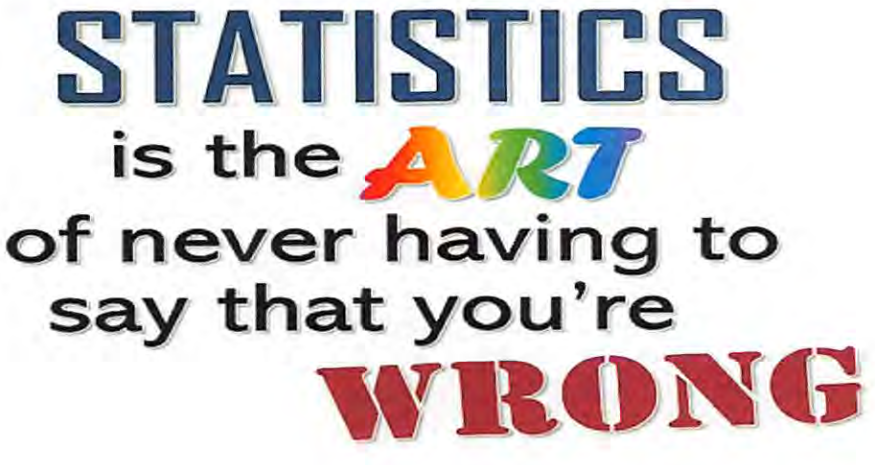
DO NOT WAIT : Please do not wait till the last minute to complete this work. This is the base foundation for the entire course and must be done thoughtfully. Pace yourself and play around with the functions of the calculator to get as familiar as possible coming into the first day.

NEAT: All work should be done neatly and organized in the packet. Please make sure you label graphs and I am able to find your answers.

DUE: This packet is due on the first day of class, please come to the first day prepared with it done. If it is turned in late, it will significantly impact your grade.

QUIZ: Be prepared to take a quiz in the first two weeks of school, this will be a good indicator of whether AP Stats is the right course for you.

Have a great summer and I look forward to having you in class in September.



Activity 1: Types of Data

There are two types of data: quantitative and qualitative (also called categorical)

Quantitative variables (think quantity) are data that take on numerical values that actually represent a measurement such as size, weight, how many, how long, score on a test, etc. For these data, it makes sense to find things like "average" or "range". For instance, it doesn't make sense to find the mean eye color because eye color is not an example of a quantitative variable. There are two types of quantitative date, **discrete** and **continuous**.

Discrete values are those such a shoe size (6, 6.5, 7, 7.5, 8, 8.5, ...), the number of cans collected by each homeroom during the Eden Student Council food drive, or class size. These variables have a finite, or countable, number of values.

Continuous variables are those such as height (60 in., 62.45 in., 63.342 in.) or how much water it takes to fill a water balloon before it bursts. These variables can assume an infinite number of values, and can assume any decimal quantity within a small range of values, even though we may choose to round the answer (as in our height). These are typically values that result from some kind of measurement, like height, weight, surface area of an orange, ERA in baseball, GPA, etc.

Qualitative (Categorical) variables are data that categorize individuals, or place them in groups. Examples are eye color, gender, year in school (junior, senior, ...). Within the qualitative group we find binary variables. A binary variable is a qualitative variable that has only two outcomes. Gender (male or female), do you have your license, do you play the piano, results of flipping a coin...)

NOTE: *Just because your variable's values are numbers, don't assume that it is quantitative.* For example, a social security number is a numerical output that is not quantitative. You are not the 117,565,487 person born in the USA. The area code is a designation to a region, not a numerical quantity. Your year in school is also not quantitative; rather it categorizes you as a 9^{th}, 10^{th}, 11^{th}, or 12^{th} grader.

Summary:

Problem set 1: Once you answer the question, determine if the data is Quantitative (discreet or continuous) or Qualitative (binary or not binary).

Question:	Answer	Type
1. In what grade did you take Algebra 1?		
2. How many DVD's do you own?		
3. How old was your father when you were born?		
4. What is your zip code?		
5. What score do you want on the AP exam?		
6. How many siblings do you have?		
7. Do you like broccoli?		
8. What is your favorite subject?		
9. What is your gender?		
10. How tall are you (in inches)?		
11. How many AP classes will you be taking this year?		
12. How many cousins do you have?		
13. How long have you lived in your current home?		
14. How far do you live from school?		

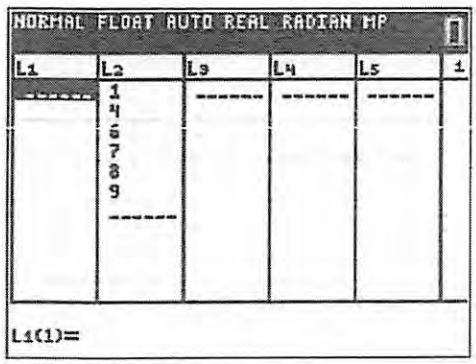
Activity 2: Data and Lists

Quantitative data can be stored in lists on the TI-84 calculator. There are several ways to create a list. From the home screen, braces can be used to define a data set, which then can be stored in one of the list names L_1 through L_6 (see figure 1.1) It seems much easier to use the calculator's editing abilities. From the STAT menu, choose edit: you may now enter the data into any of the lists (figure 1.2)

Figure 1.1:

$\{1,4,6,7,8,9\} \rightarrow L_2$

Figure 1.2:



In either case, new lists can be created from existing lists, such as $L_2 + 5$ (figure 1.3 and 1.4)

Make sure when you enter the command that you are on the L_3 icon and not within the list of numbers.

Figure 1.3

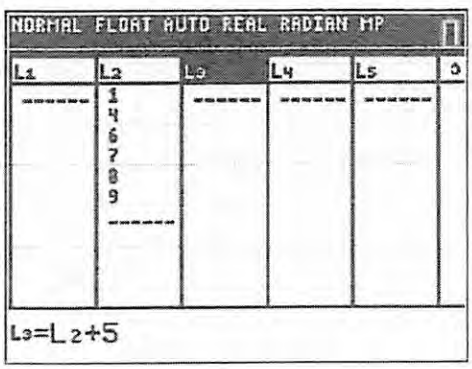
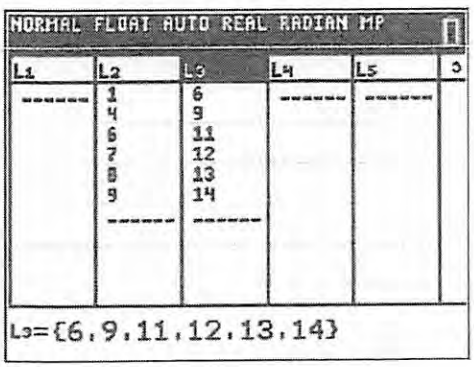


Figure 1.4



Lists may also be given their own names and will be retained in memory until deleted. This is particularly useful for data that will be used repeatedly.

Example 1: Create a named list for the following set of running speeds in mph for various animals:

Cheetah	70	Warthog	30
Lion	50	Cat	30
Coyote	43	Man	27.89
Hyena	40	Pig	11
Greyhound	39.35	Tortoise	.17
Rabbit	35	Snail	.03

Source: 1996 Information Please Almanac

Procedure: to create a named list go to the list editor and move to the right past L_6 . A "name" prompt will appear and the list name can be typed (figure 1.5). The values can be entered in the usual way (figure 1.6)

Figure 1.5

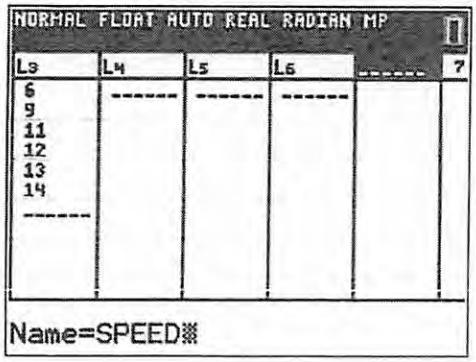
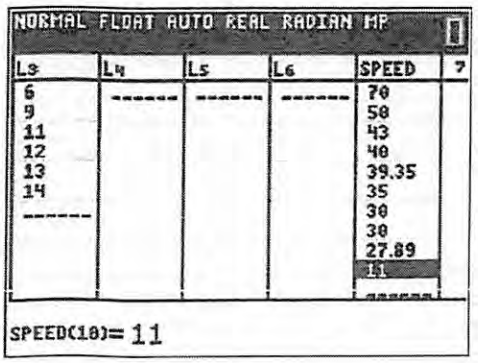


Figure 1.6



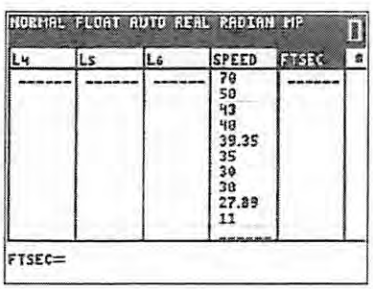
Example 2: Create a new list showing the speeds in feet per second.

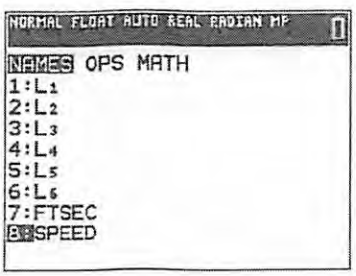
Procedure: New lists can be created from named lists on either the home screen or in the list editor. On the home screen, the speeds in mph can all be converted to ft/sec and stored in a list named FTSEC by a single command (figure 1.7). In order for the TI-84 to distinguish a user defined list name from other symbols it is necessary to preface a list name with a special character L that is located in the LIST: OPS sub-menu. The L character may also be found in the CATALOG. First create a list named FTSEC using the steps outlined above.

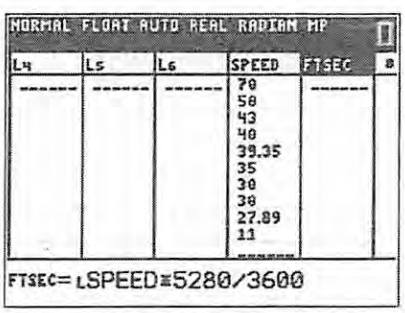
From the home screen:

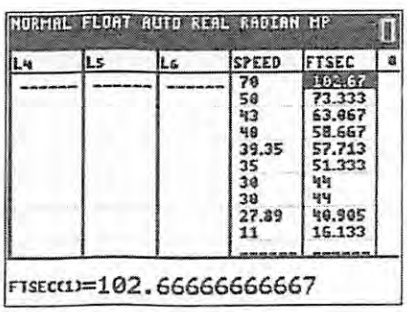
Figure 1.7: $_L SPEED * 5280 / 3600 \rightarrow {}_L FTSEC$

From the list editor: Highlight the list name icon FTSEC, then go to the LIST: Names (2^{nd} STAT) , pick the list named SPEED, multiply by 5280 and divide by 3600, enter.









Problem set 2

Exercises:

1. Create a list L1 using {4,7,9,11,14,17,20}

   From that, create new lists:

   A. $L_1 - 7$: _____
   B. 2*L1: _____
   C. L_1^2 : _____
   D. $\sqrt{L_1}$: _____
   E. $\ln L_1$: _____

2. Create a list PLANM showing the mean distance frm the sun in millions of miles for each planet. Create a new list PLANK showing the mean distance in millions of kilometers. (Use the internet to find the miles:kilometers conversion)

Planet	Mean distance in millions of miles	Mean distance in millions of kilometers
Mercury	36.0	
Venus	67.24	
Earth	92.9	
Mars	141.71	
Jupiter	483.88	
Saturn	887.14	
Uranus	1783.98	
Neptune	2796.46	

Activity 3: Numerical Descriptions of Quantitative Data

There are two categories of numbers that are used to describe a set of data: measures of center and measures of spread.

Measures of Center:

1. The **Mean** is the average number. It is the sum of all the data values divided by the number (n) of values.

Example: {4, 36, 10, 22, 9} mean = $\bar{X} = \frac{\sum X_i}{n} = \frac{4+36+10+22+9}{5} = 16.2$

2. The **median** is the value that separates the bottom 50% of data from the top 50% of data. It is the middle element of an ordered set of data that is odd in number. It is the average of the two middle elements of an ordered set of data that is even in number.

Example: {4, 9, 10, 22, 36} median is 10

{4, 9, 10, 22, 36, 43} median is $\frac{10+22}{2} = 16$

3. The **mode** is the value that occurs most often in a set of data. If the data occurs with the same frequency, then there is no mode. If two (or more) values occur the most then they are both the mode. We call this bimodal.

Measures of Spread:

1. The **range** is measure of spread of the entire data. It is calculated by subtracting the minimum value from the maximum value.

Ex: {4, 36, 10, 22, 9, 43} = {4, 9, 10, 22, 36, 43} range = 43 – 4 = 39

2. The **interquartile range (IQR)** is a measure of the spread of the middle 50% of the data. It is calculated by subtracting the 25th percentile (Q1) from the 75th percentile (Q3). Q1 is the median of the lower half of the data. It separates the bottom 25% of values from the top 75% of values. Q3 is the median of the upper half of the data. It separates the top 25% of values from the bottom 75% of values. In neither of these cases is the median considered in the top half or the bottom half of the data.

Ex: {4, 9, 10, | 22, 36, 43}

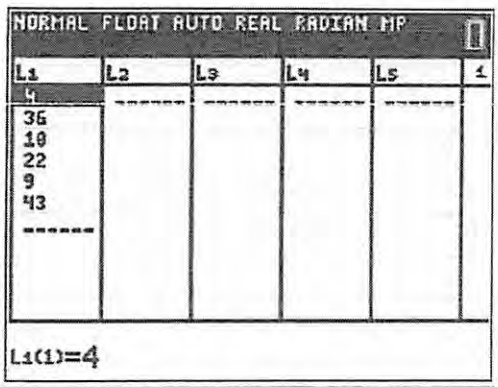
Q1 Q3 IQR = 36 – 9 = 27

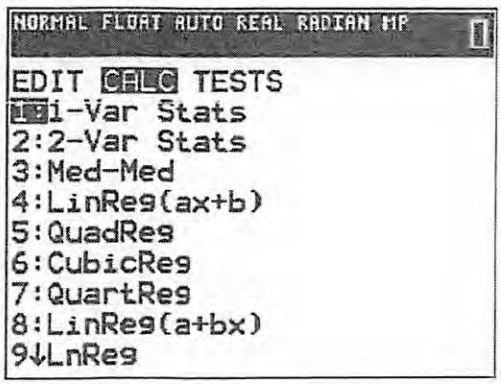
3. The **standard deviation** is the measure of the spread around the mean. It is calculated using the formula:

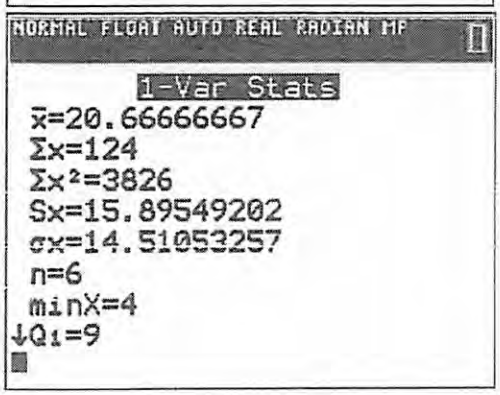
$$s = \sqrt{\frac{\Sigma\left(x_i - \bar{X}\right)^2}{n-1}}$$

Luckily, we don't calculate the standard deviation by hand. It is easier to put the data into List 1 and calculate the 1-VAR STATS.

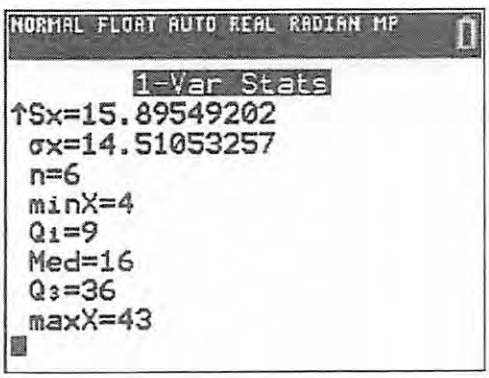
To see statistical results including the quartiles, mean and standard deviation, use STAT CALC 1:1-Var Stats







If you arrow down you will see the rest of the stats.



Problem set 3:

Here is a list of parents' ages at the time their sons were born

Dad:	41	27	23	31	30	33	26	32	43	25	34	27	25
	34	27	26	28	32	32	35	27	33	34	34	34	35
Mom:	39	26	23	30	28	33	23	32	38	23	35	24	24
	33	24	32	23	30	24	29	34	35	26	31	23	37

Enter these two lists into your calculator and use the 1-Var Stat option to calculate the following:

1. Data for Dad: mean _____ median _____ Which is larger? _____

2. Data for Mom: mean _____ median _____ Which is larger? _____

3. Now compare the two means. Which is larger? _____

4. Is this what you expected? _____ Explain why or why not. _____

5. Calculate the standard deviations for both sets of data: Dad _____ Mom _____

Why might these be different? _____

6. Find Q1 and Q3 and the IQR for Dad: Q1 _____ Q3 _____ IQR _____

Mom: Q1 _____ Q3 _____ IQR _____

7. A company has two machines that fill cans of soft drinks. Samples from each machine show the following number of ounces per can:

Machine A: 11.1, 12.0, 11.4, 12.1, 11.7, 11.5, 12.2, 11.4, 11.3, 11.9

Machine B: 10.9, 12.4, 12.7, 11.8, 12.3, 11.9, 12.0, 12.5, 12.7, 11.6

Find the mean and standard deviation for both machines.

$A_{\bar{x}}$ = _____ A_{sd} = _____ $B_{\bar{x}}$ = _____ B_{sd} = _____

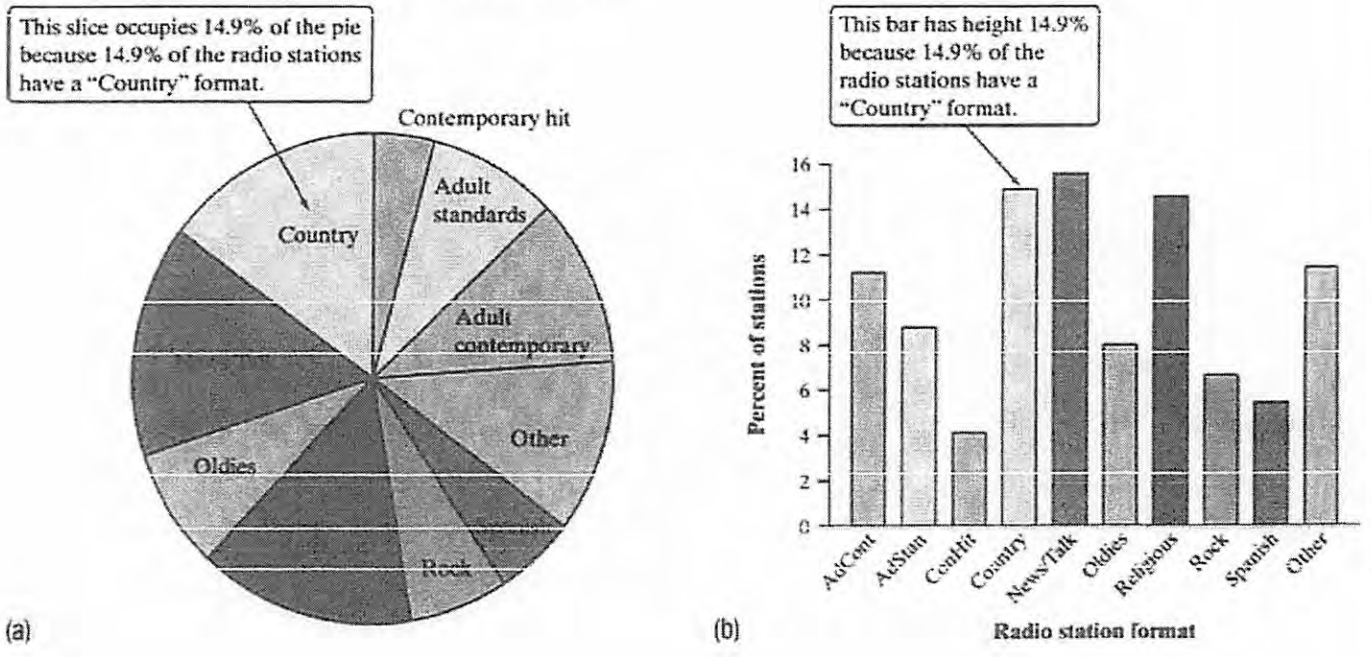
8. Using your answer to #7, explain which machine is "better" at filling soft drink cans.

There are two basic graphical techniques used for qualitative data as described below:

1. Bar graph – one axis is the values of your variable (order does not matter) and the other axis is the frequency or relative frequency (percentages). The bars do not touch.

2. Pie chart – the whole group of values is shown as a circle. Each "piece of the pie" corresponds to the relative frequency of the values. To determine the central angle that forms each "piece" use the following: central angle= 360 (relative frequency)

Radio Station formats: pie graph vs. bar graph



You should have extensive practice with these techniques so we won't cover them in this packet.

Activity 4: Graphical Displays of Univariate Data (One variable)

Histograms: A set of quantitative data is often separated into groups or intervals. For example, test scores may be separated into 90-99, 80-89, 70-79, and so on. The frequency refers to the count for each interval. A histogram is a visual representation of the frequencies plotted against the interval.

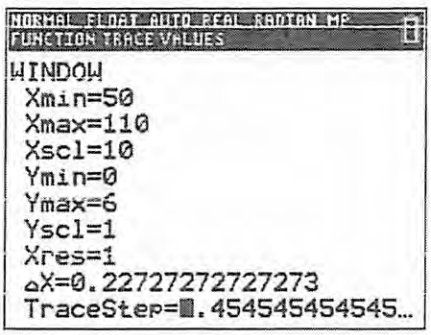
Exercise 1: Enter the following test scores into List 3 and create a histogram using intervals of width 10:

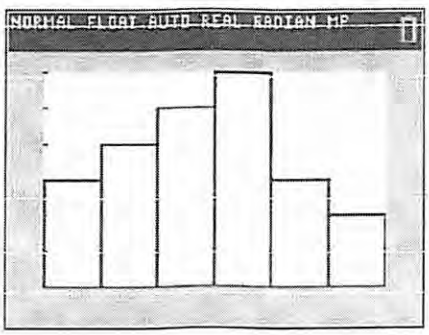
100, 100, 98, 95, 92, 88, 85, 84, 84, 81, 80, 77, 75, 72, 70, 70, 66, 65, 63, 60, 59, 55, 50

Procedure: First enter the scores into L3. You must now turn on your STATPLOT. The 2nd function of the Y= will get you to the menu.

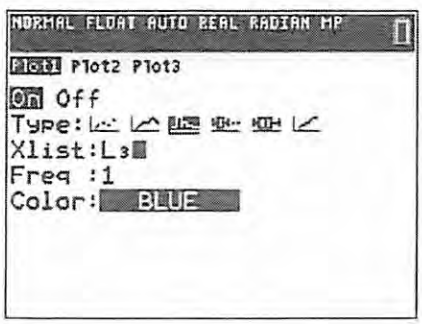
Choose one of the Plots by pressing 1, 2, or 3. Now activate the plot by selecting ON. At the TYPE prompt choose the third icon for histogram. At the X LIST, make sure it says L3 (2nd function on the numbers 1 to 6). Frequency should be 1.

To control the width of the bars (called the bin width), set up the window manually. Xscl=10

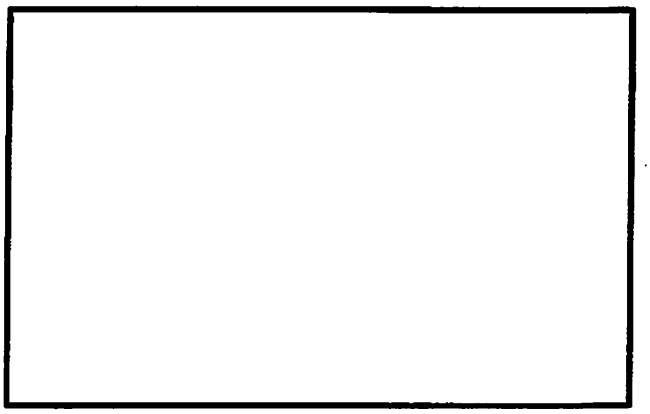




You may now TRACE the graph to see the intervals and their frequencies. When drawing a histogram, make sure both axes are labeled and have appropriate scales.



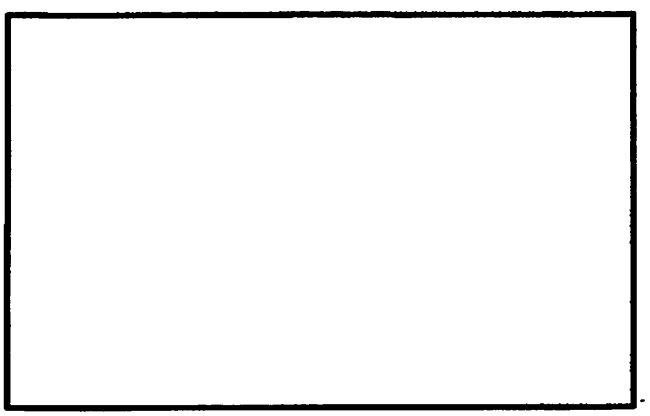
Now go back to the WINDOW for this example and change the Xscl to 5 and GRAPH. Draw the histogram below and describe how this one differs from the first histogram. Which histogram would best describe the data?



Go back and change the Xscl to 15 then graph. Describe how this histogram differs from the first two. Aagain, state which histogram would best describe the data.

From these two problems, what conclusion can you draw about how the Xscl affects the histogram?

Create a histogram of the list PLANM (Activity 2: exercise 2) which shows the average distance from the sun for the planets of our solar system.



What would be the best setting for the Xscl with this data and why?

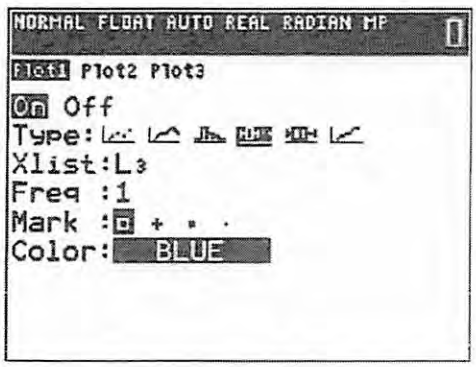
Example 2: **Boxplots**

A box plot is a graphical display for a set of quantitative data that only displays the five number summary:

minimum, Q1, median, Q3, maximum

Using the same data as before (example 1 - test scores), create a boxplot below. Make sure your axis has a scale and label.

In the STATPLOT menu, you will notice two boxplot options. The first one allows for possible outliers. I recommend you always use this one.



Suppose another class receives these test scores:

94, 93, 90, 83, 82, 82, 77, 76, 75, 70, 64, 60, 53, 52, 48, 44, 40

Create two box and whiskers plots on the same set of axes to compare their performance with the original group.

Write a few sentences comparing the medians and the spreads of each class.

Refer to the company that has two machines that fill soda cans (Activity 3: exercise 7). Create a box plot for each machine. Sketch one above the other.

Find the range for Machine A _____ Machine B _____

Find the median for Machine A _____ Machine B _____

Which machine is more consistent? _____ Why? _____

_____

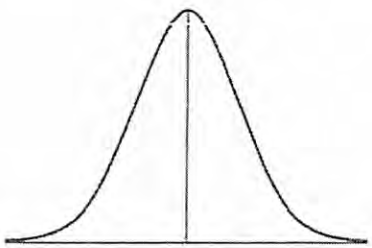
Activity 5: Assessing the Shape of a Graph

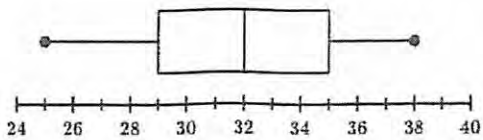
When describing a set of data we look at the following features:

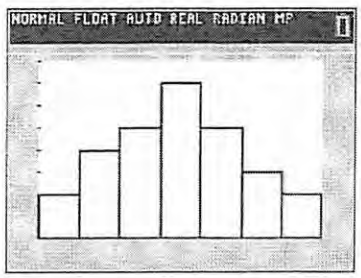
Shape center spread clusters and gaps outliers

We have several terms that we use to describe the shape but this packet will concentrate on only two: symmetric and skewed.

You can tell if a graph is symmetric if a vertical line in the center divides the graph into two fairly congruent shapes. The following sets of data can be described as symmetric:

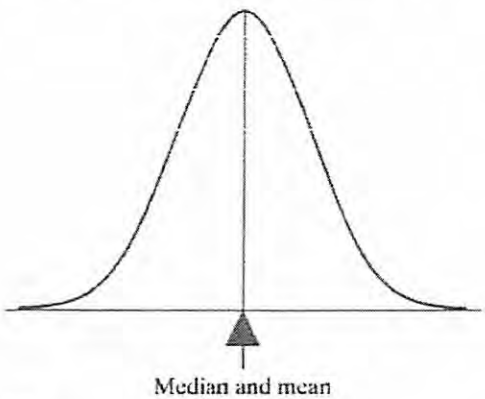




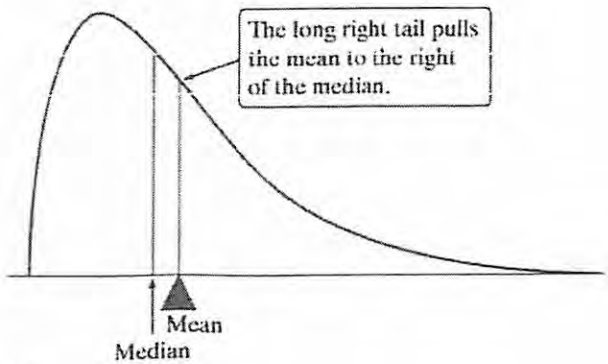


The mean and the median are approximately the same in a symmetric graph. Figure (a) below.

One can tell if a graph is skewed if the graph has a big clump of data on either the left or the right with a tendency to get flatter on the opposite side.

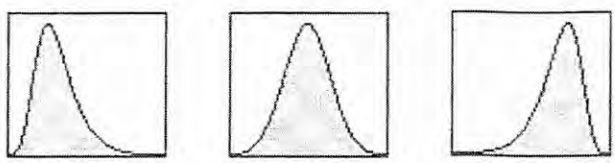


(a)



(b)

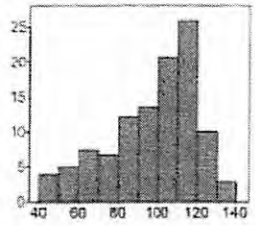
The distribution in figure (b) is skewed right. The mean gets pulled toward the tail, which is the direction of the skewness. So the mean > median



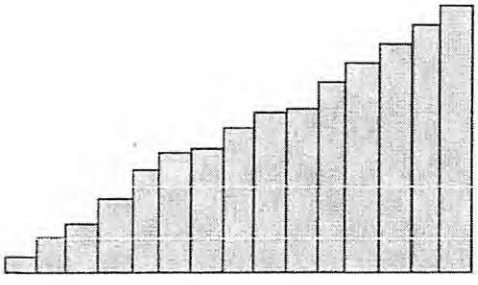
The mean is always pulled closer toward the tail.

Exercise 1. Determine if the distribution is symmetric, skewed left or skewed right. Then determine if the mean or the median is the larger value. If they are equal, state so.

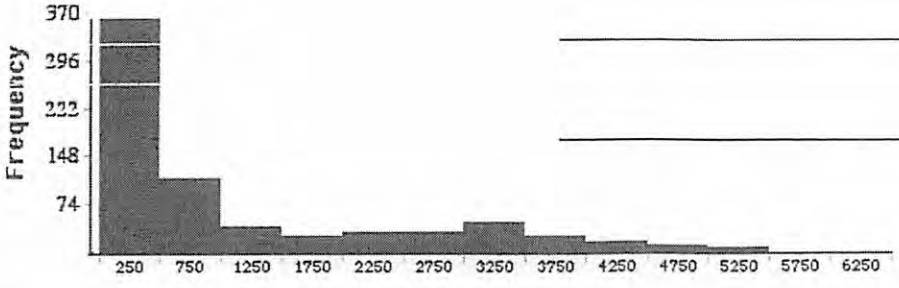
1.



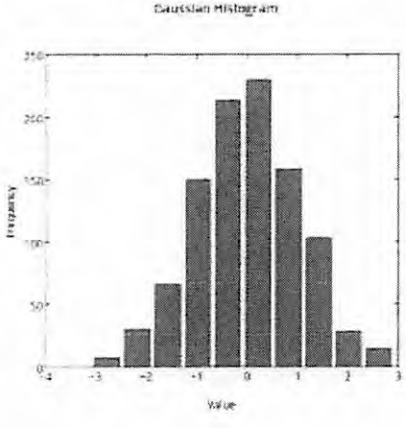
2.



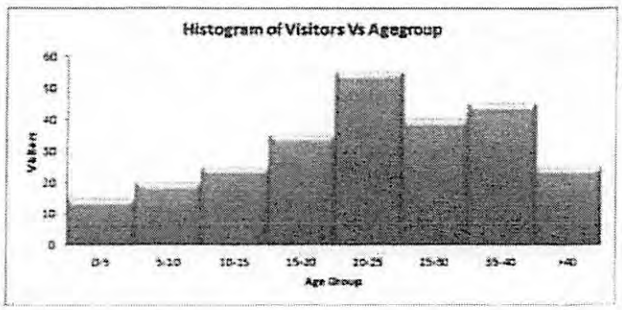
3.



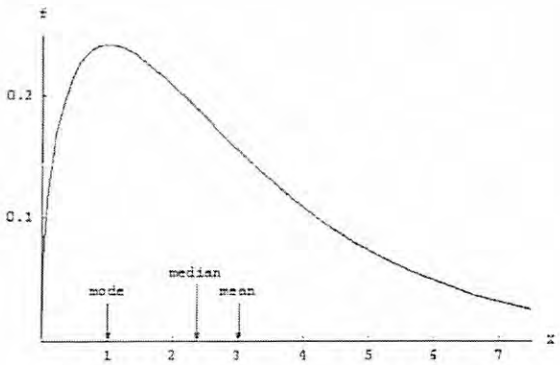
4.



5.


Histogram of Visitors Vs Agegroup
Visitors
Age Group

6.


f
0.2
0.1
mode
median
mean
1
2
3
4
5
6
7
x

7.

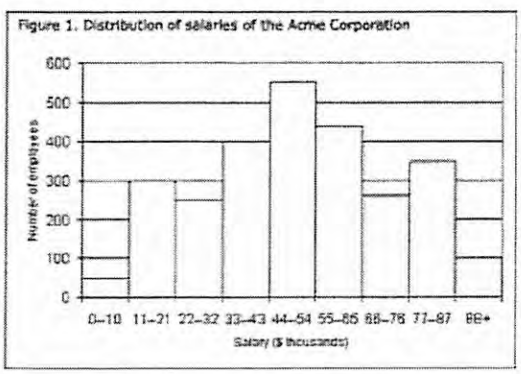
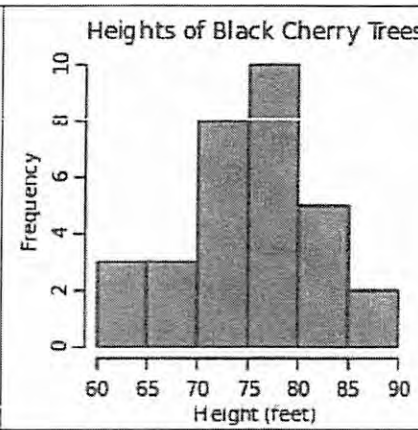
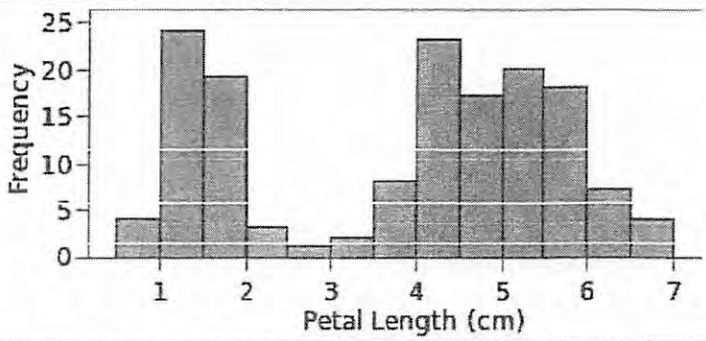
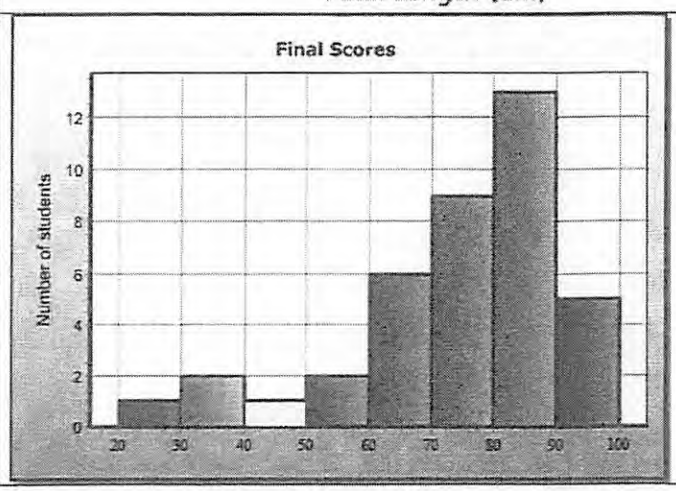
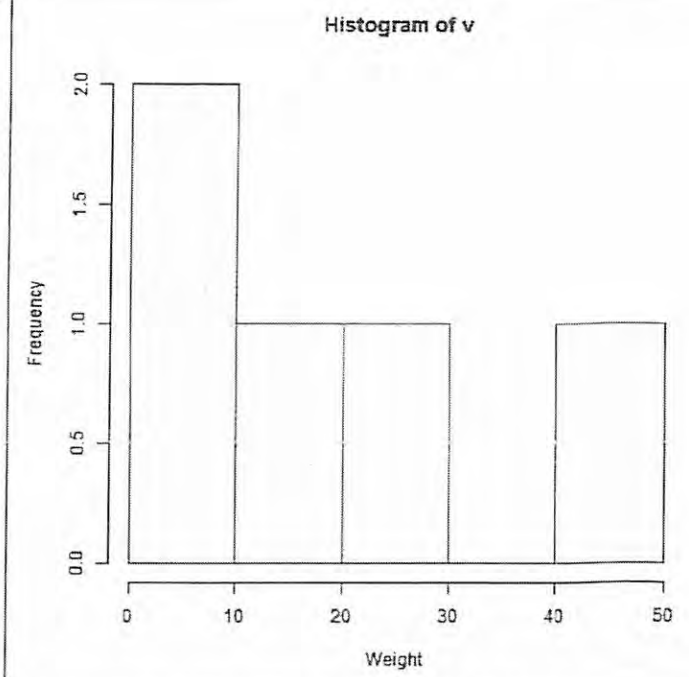
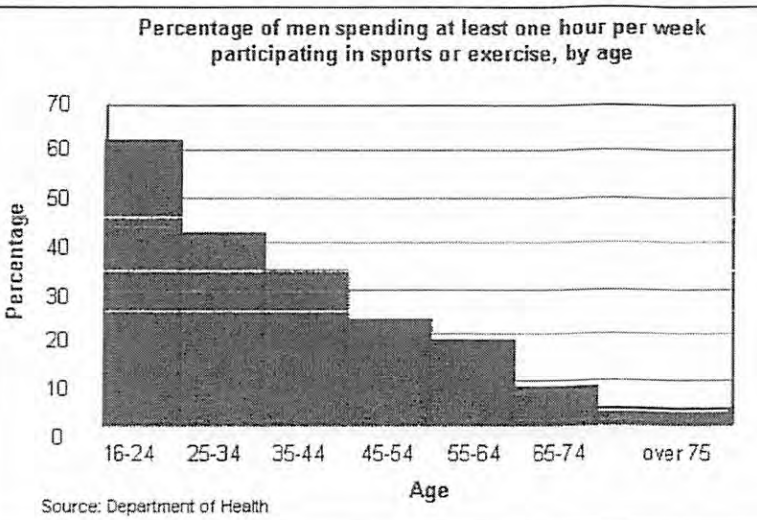


Figure 1. Distribution of salaries of the Acme Corporation
600
500
400
300
200
100
0
Number of employees
0–10
11–21
22–32
33–43
44–54
55–65
66–76
77–87
88+
Salary ($ thousands)

8. For the following graphs, find the shape, center (use the median), and spread (use the range). If there are any other notable features evident in the graph (clusters, gaps or outliers), then say where they are. Otherwise, do not comment on them.

Note: To find the center of these graphs, use the frequencies found on the y-axis. Count how many are in each bar. If the bar has an interval, use the midpoint. Add these up and divide by 2. This tells you where the median is located. Which bar is this value located in?

1		Shape _____ Center _____ Spread _____ Clusters? _____
2		Shape _____ Center _____ Spread _____ Clusters? _____
3		Shape _____ Center _____ Spread _____ Clusters? _____

4		Shape _____ Center _____ Spread _____ Clusters? _____
5		Shape _____ Center _____ Spread _____ Clusters? _____

Good job! You have completed this packet. Please have it ready to turn in on the first day of class. If you have any questions about anything in this packet, please write them down so you can ask them on day 1.

Ms. Braillard