Data About Us Notes
Problem 1.1

Everyone has a name. Did you ever stop to think about how many letters are in your name?

What do you think is the typical number of letters in the full names (first and last names) of students in your class.

Look at the list of names below:

Name	# of Letters	Name	# of Letters
Al Pine	6	Phil Upp	7
Justin Case	10	Moe Tell	7
Nick O'Time	9	Clay Potts	9
Penny Pincher	12	Ken Knee	7
Kowa Bunga	9	Stu Dent	7
Joe Kerr	7	Penn Head	8

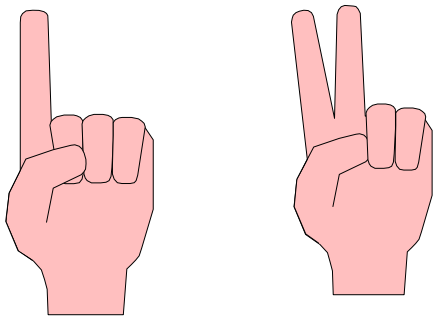
The following are some statements that you could make about the collected data.

1. *The fewest letters in a name was six.*

2. *The most letters in a name was twelve.*

3. *The range of numbers was from 6 to 12.*

4. *The most frequent number of letters in a name is seven.*



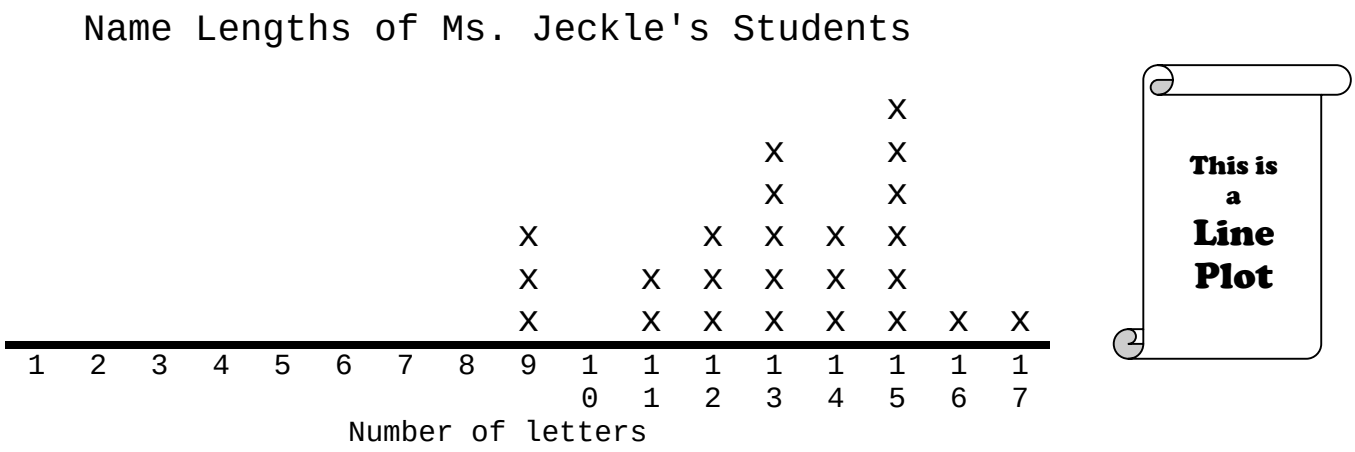
Data About Us Notes
Problem 1.2

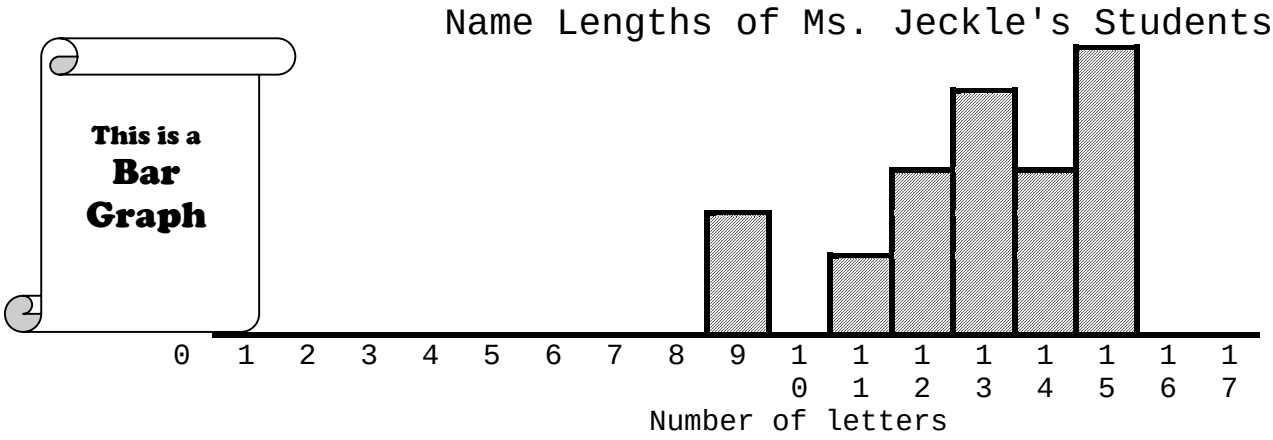
Definitions:

line plot - a way to organize data on a number line by using an "x" as a frequency tally

bar graph - a way to display data where a bar shows the frequency of a particular event

These are graphs of the name lengths of students from Ms. Jeckle's Class.





A. Write some statements about the name lengths for students in Ms. Jeckle's class.

/ *There is one peak at a name length of 15 letters.*

/ *The length of the names ranges from 9 letters to 17 letters.*

/ *Only two of the names are longer than 15 letters.*

/ *Most of the name lengths cluster in the interval of 12-15 letters.*

B. In what ways are the two graphs alike? In what ways are they different?

/ *Both graphs have titled and labeled axis.*

/ *They have similar horizontal scales, which show the possible name lengths.*

/ *The line plot does not have a vertical scale. The frequency of a particular name is indicated by an "x" above that number.*

/ The bar graph has a vertical scale, which is needed to determine the frequencies based on the height of each bar.

Data About Us Notes
Problem 1.3

Definitions:
mode - the value that occurs most frequently in a set of data
range - the spread of data values from the lowest value to the highest value

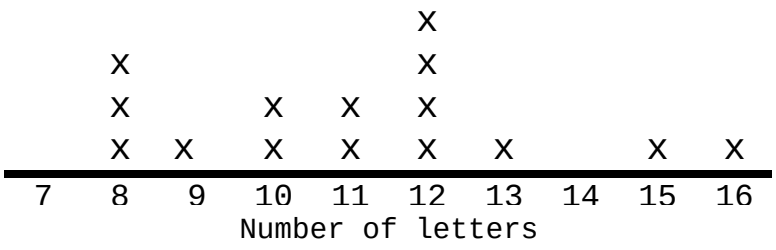
Suppose there are 15 students in a class. The mode of the name lengths for the class is 12 letters, and the range is from 8 letters to 16 letters.

A. Determine a set of name lengths that has this range and mode.

possible data set:

8 8 8 9 10 10 11 11 12 12 12 12 13 15 16

B. Make a line plot to display your data.



C. You can use your line plot to describe the shape of your data.

- *The data may be displayed in a bell-shape. This would mean that there were more "X's" in the center of the line plot than at the ends.*
- *The data may be spread out so that there are two or more clusters of "X's."*
- *The data may be spread out equally so there each number has about the same amount of "X's."*
- *The data may be displayed so that all of the "X's" are grouped together at one end of the line plot.*

Data About Us Notes
Problem 1.4

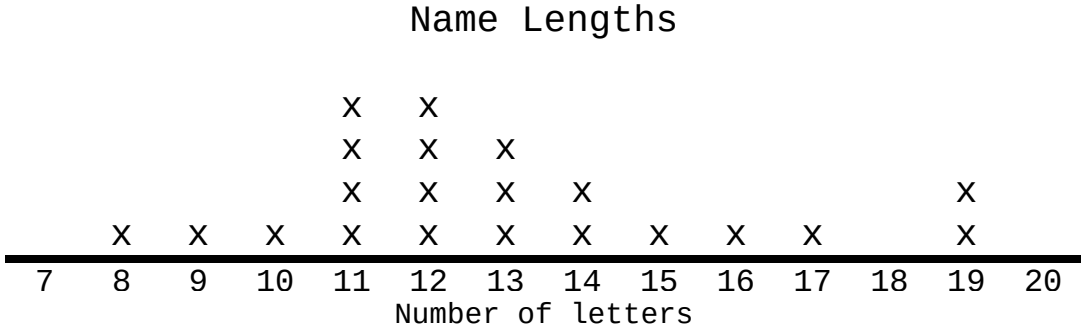
Definitions:

median - the value that is in the middle of the data set; it cuts the data in half

The data set and line plot below show name length data from a middle school class in Michigan. Notice that this data has two modes, 11 and 12. The range of the data is from 8 letters to 19 letters.

Name length in letters:

19 11 15 14 13 12 19 11 12 13 11
13 16 8 11 10 9 17 14 12 12



A. Suppose you were to write all of the number of name lengths on a strip of paper like the one below, and then fold the strip of paper in half.

8	9	10	11	11	11	11	12	12	12	12	13	13	13	14	14	15	16	17	19	19

Where would the crease land?

The crease lands **on** *12.*

How many numbers are to the left of the crease?

There are 10 values to the left of the crease.

How many numbers are to the right of the crease?

There are 10 values to the right of the crease.

B. Suppose there was a new student that had a name length of 19 letters. If you add her name to the list, it would look like the one below.

8	9	10	11	11	11	11	12	12	12	12	13	13	13	14	14	15	16	17	19	19	19

Where is the crease?

The crease lands **between** *12 and 13.*

How many numbers are to the left of the crease?

There are 11 values to the left of the crease.

How many numbers are to the right of the crease?

There are 11 values to the right of the crease.

Problem 1.4 Follow-Up

When there is an odd number of data values:

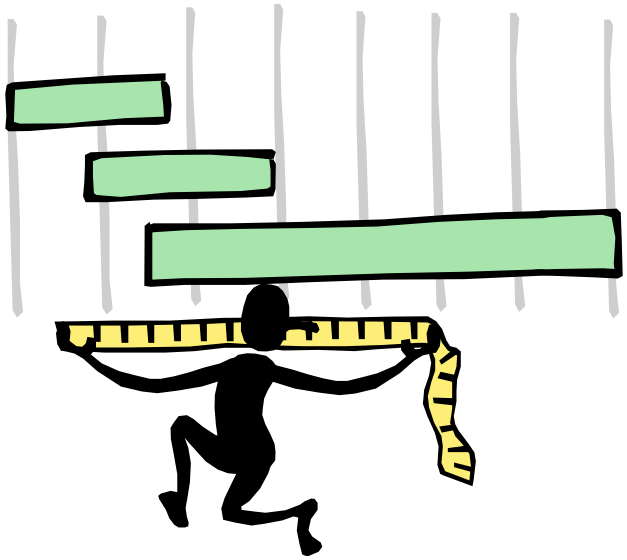
The first strip of paper has 21 data values. When you folded the strip, **the crease was on the number 12**. There were ten values to the left of the crease and ten values to the right of the crease. We say that 12 is the **median** of the data set.

When there is an even number of data values:

The second strip you made has 22 values. When you folded this strip, **the crease landed between 12 and 13**. There were eleven values to the left of the crease and eleven values to the right of the crease. When a data set has an even number of values, the median is the value halfway between the two middle values. For this data set, the median is 12.2, the number halfway between 12 and 13.

Measures of central tendency:

Giving the median of a set of data is one way to describe what is typical about the data. Like the mode, the median is a type of average. The median and the mode are sometimes referred to as **measures of center** or **measure of central tendency**. You can see this is a very appropriate description for the median since it is the center of the data.

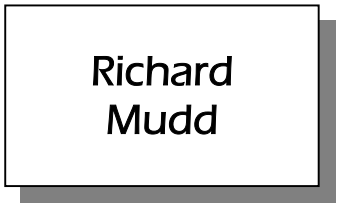


Data About Us Notes
Problem 1.5

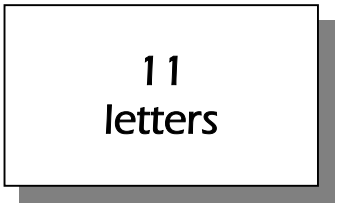
- **What happens to the median when you add values to or remove values from a set of data?**
- **Does adding a value that is much larger or much smaller than the rest of the data values have a greater effect on the median than adding a value that is closer to the other values?**

Suppose you were to write each of the names listed below on an index card. On the back of each card, write number of letters in a name. Order the cards from shortest name to longest name and find the median of the data.

Name	Letters
Thomas Petes	*11*
Michelle Hughes	*14*
Shoshana White	*13*
Deborah Locke	*12*
Tonya Stewart	*12*
Richard Mudd	*11*
Tony Tung	*9*
Janice Vick	*10*
Bobby King	*9*
Kathleen Boylan	*14*



front



back

Experiment with the cards and try to perform each task described below.

A. Remove two names without changing the median.
- *Remove the lowest value, 8, and the highest value, 14; the median remains 11.5.*
- *Remove the two middle values; the median remains 11.5.*

B. Remove two names so the median increases.
- *Remove 8 and 9; the median increases to 12.*
- *Remove 10 and 11; the median increases to 12.*

C. Remove two names so the median decreases.
- *Remove both 12's; the median decreases to 11.*
- *Remove both 14's; the median decreases to 11.*

D. Add two new names so the median increases.
- *Add 16 and 17; the median increases to 12.*
- *Add 900 and 1000; the median increases to 12.*

E. Add two new names so the median decreases.
- *Add 4 and 5; the median decreases to11.*
- *Add 11 and 11; the median decreases to 11.*

F. Add two new means without changing the median.
- *Add 11 and 12; the median remains 11.5.*
- *Add 1 and 2000; the median remains 11.5*

Data About Us Notes
Problem 3.1

Definitions:

stem-and-leaf plot - a way to display data that shows the distribution and actual values

You have looked at line plots and bar graphs as ways to display data and look at the shape of the distribution. Another way to display data with a large range of numbers is by using a stem-and-leaf plot. The stem-and-leaf plot below uses the data on the back of page.

Step 1: Look at all the values in your data set. Look at the tens place and determine the range of values. Since or range is from 5 to 60, we will use the following numbers: 0, 1, 2, 3, 4, 5, and 6.

Step 2: Make a vertical line and then make a vertical list of the numbers to the left of the line.

Step 3: Add the "leaves." These are the ones values.

0	8	8	5	5	5	6											
1	5	5	5	5	9	5	5	7	5	0	5	0	5	1	7	0	0
2	2	5	0	5	0	0	0	0	0	1							
3	0	5	0	0	5												
4																	
5	0																
6	0																

Step 4: Redraw the plot so that the values for each set of tens is arranged from smallest to largest.

Step 5: Add a key and a title.

Travel Times to School (minutes)

0	5	5	5	6	8	8											
1	0	0	0	0	1	5	5	5	5	5	5	5	5	5	7	7	9
2	0	0	0	0	0	0	1	2	5	5							
3	0	0	0	5	5												
4																	
5	0																
6	0																

2 | 5 means 25 minutes

A. Which students probably get to sleep the latest in the morning? Why?
Students who have the shortest travel times probably sleep the latest. This would be students in the 0-9 minute interval or the 10-19 minute interval.

B. Which students get up earliest? Why?
Students who have longer travel times probably get up earlier. These would be students in the 20-29 or 30-39 minute intervals.

C. What is the typical time it takes for these students to travel to school?
If you use the median, it is about 16 minutes.

Student's Initials	Time (minutes)	Distance (Miles)	Mode of travel
DB	60	4.50	bus
DD	15	2.00	bus
CC	30	2.00	bus
SE	15	0.75	car
AE	15	1.00	bus
FH	35	2.50	bus
CL	15	1.00	bus
LM	22	2.00	bus
ON	25	1.5	bus
NP	20	1.5	bus
AP	25	1.25	bus
AP	19	2.25	bus
HCP	15	1.50	bus
KR	8	0.25	walking
NS	8	1.25	car
LS	5	0.50	bus
AT	20	2.75	bus
JW	15	1.50	bus
DW	17	2.50	bus
SW	15	2.00	car
NW	10	0.50	walking
JW	20	0.50	walking
CW	15	2.25	bus
BA	30	3.00	bus
JB	20	2.50	bus
AB	50	4.00	bus
BB	30	4.75	bus
MB	20	2.00	bus
RC	10	1.25	bus
CD	5	0.25	walking
ME	5	0.50	bus
CF	20	1.75	bus
KG	15	1.75	bus
TH	11	1.50	bus
EL	6	1.00	car
KLD	35	0.75	bus
MN	17	4.50	bus
JO	10	3.00	car
RP	21	1.50	bus
ER	10	1.00	bus

Data About Us Notes
Problem 3.2

Definitions:

outliers - values that stand out in a set of data (very large or very small values)

You can compare two different sets of data by making a back-to-back stem-and-leaf plot. The one below shows the number of jumping jacks that students recorded in two different classes.

Number of Jumps

Ms. R's class								*Mr. K's class*								
8	7	7	7	5	1	1	0	1	1	2	3	4	5	8	8
				6	1	1	1	0	7						
	9	7	6	3	0	0	2	3	7	8					
					5	3	3	0	3	5					
					5	0	4	2	7	8					
							5	0	2	3					
						2	6	0	8						
							7								
				9	8	0	8								
				6	3	1	9								
							10	2	4						
						3	11								
							12								
							13								
							14								
							15	1							
							16	0	0						
							17								
							18								
							19								
							20								
							21								
							22								
							23								
							24								
			Key:				25								
							26								
7 | 3 | 0			*means 37 jumps for Mr. R's class and 30 jumps for Mr. K's class.*				27								
							28								
							29								
							30	0							

Which class did better overall in the jump-rope activity?

One possible answer is:

The median number of jumps for Ms. R's class was 26.5, with a range of 1 to 113 jumps while the median number of jumps for Mr. K's class was 34, with a range of 1 to 300 jumps.

In Mr. K's class, there are some very large numbers of jumps. For example, one student jumped 151 times, and another student jumped 300 times. We call these data outliers, or numbers that stand out in a set of data.

Data About Us Notes
Problem 4.1

Definitions:

coordinate graph - a way to display two related numbers
x-axis - the horizontal axis on a coordinate grid
y-axis - the vertical axis on a coordinate grid

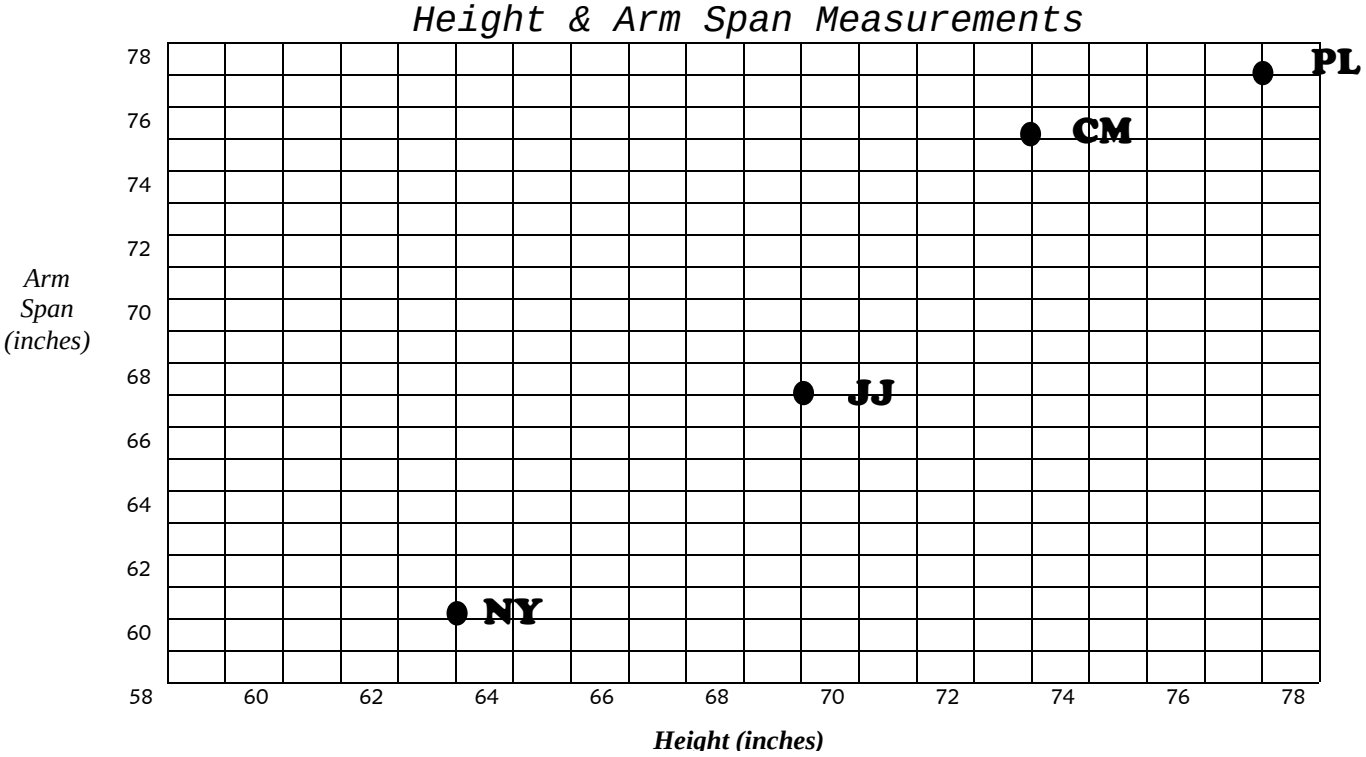
Are a person's height and arm span related?

To answer the question, let's analyze the data in the table. Each set of data also has a coordinate pair. These are used to graph points on a coordinate grid.

Height and Arm Span Measurements

Initials	Height (inches)	Arm Span (inches)	Coordinate Pairs
NY	63	60	(63, 60)
JJ	69	67	(69, 67)
CM	73	75	(73, 75)
PL	77	77	(77. 77)
BP	64	65	(64, 65)
AS	67	64	(67,64)
KR	72	72	(72,72)

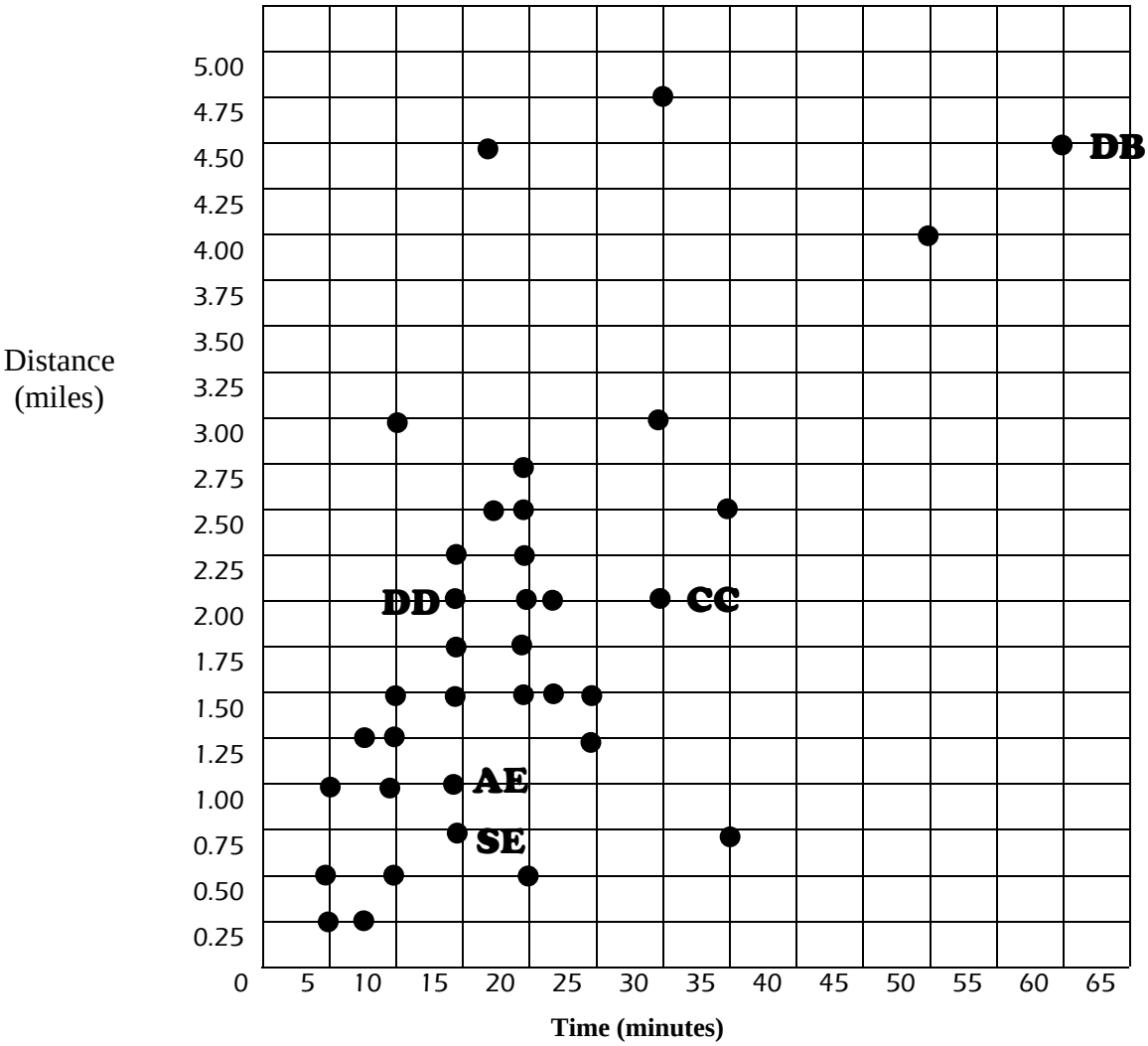
You can graph the data on a coordinate grid. The horizontal axis is the x-axis and the vertical axis is the y-axis.



Can you graph the remaining points?

Data About Us Notes
Problem 4.2

In Investigation 3, you made a stem and leaf plot to show data about travel times to school. Using a coordinate graph, you can show both travel time and distance from home to school. The graph below shows the travel times and distances used in Investigation 3. The location of the first five students on the list have been identified by their initials.



1. If you look at the points (17, 4.50) and (60, 4.50), what information can these points give you?

   Each student is traveling the same distance (4.5 miles). It takes the first student only 17 minutes to get to school, but it takes an hour for the second student.

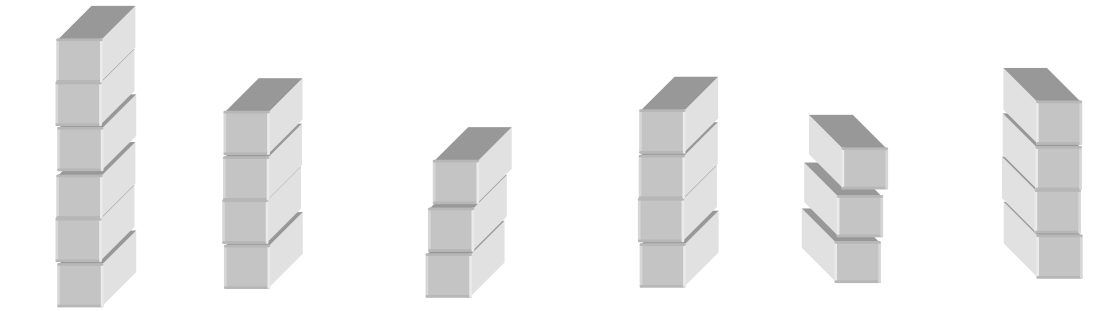
Data About Us Notes
Problem 5.2

Definitions:
mean - the average of a set of data values

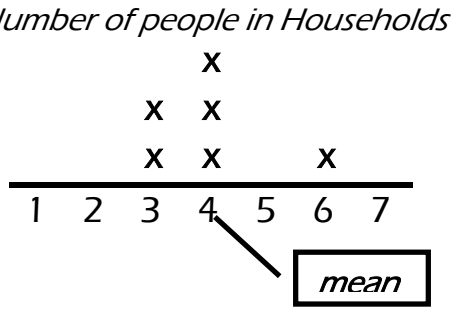
The following data show the number of people in the households of six different students.

name	# of people in household
Geoffrey	6
Betty	4
Brendan	3
Oprah	4
Reilly	3
Tonisha	4

A. Make a set of cube towers to show the size of each household.



B. Make a line plot of the data.



C. How many people, altogether, are in the six households?
6+4+3+4+3+4=24 people

D. What is the mean number of people in the six households?
24 ÷ 6 = 4 people per house

Data About Us Notes
Problem 5.5

A group of middle school students was asked this question:

How many movies did you watch last month?

Here are a table and stem plot of the data:

Student	# of movies
Joel	*15*
Tonya	*16*
Rachel	*5*
Lawrence	*18*
Meela	*3*
Leah	*6*
Beth	*7*
Mickey	*6*
Bhavana	*3*
Josh	*11*

Movies watched

0	*3 3 5 6 6 7*
1	*1 5 6 8*
2	

key

1 | *5* means 15 movies

A. The total number of students is *10*.
The total number of movies watched is *90*.
The mean number of movies watched is *9*.

B. A new value is added for Lucia, who watched 42 movies last month. This value is an outlier. How does the stem plot change when this value is added? What is the new mean? Compare the mean from part A to the mean after this value is added. What do you notice?

0	*3 3 5 6 6*
1	*1 5 6 8*
2	
3	
4	*2*

The new mean is 12 movies. It is larger than the mean for the data in part A.

C. A new value is added Tamara, who was at home last month with a broken leg. She watched 96 movies. What is the mean of the data now? Compare the names you found in parts A and B with the new mean. What did you notice and why?

The new mean is 19 movies. The value of 96 made the mean larger than either of the two earlier means.

0	*3 3 5 6 6*
1	*1 5 6 8*
2	
3	
4	*2*
5	
6	
7	
8	
9	*6*

D. Data for eight more students are added:

Student	# of movies	Student	# of movies
Tommy	*5*	*Robbie*	*4*
Alexandra	*5*	*Ana*	*4*
Kesh	*5*	*Alisha*	*2*
Kirsten	*5*	*Brian*	*2*

These data are not outliers, but now there are several students who watched only a few movies in one month. What is the mean of the data now? Compare the means you found in parts A, B, and C with this new mean. What do you notice and why?

0	*2 2 3 3 4 4 5 5 5 5 5 6 6 7*
1	*1 5 6 8*
2	
3	
4	*2*
5	
6	
7	
8	
9	*6*

The new mean is 13 movies. It has decreased from the mean of 19 movies in part C. However, the two higher values of 42 and 96 movies still influence where the mean occurs

Follow-Up

1. What happens to the mean when you add one or mover values that are much larger than the values in the original data set? Why does this happen?

   Adding unusual values to the data set can greatly affect the mean. Adding very large numbers pulls the mean up.

2. What happens to the mean when you add a number of values that are clumped with the smaller values in the original data set? Why does this happen?

   Adding several small numbers pull the mean down.