

Mathematics

8.12 The student uses statistical procedures to describe data. The student is expected to:

- (A) select the appropriate measure of central tendency or range to describe a set of data and justify the choice for a particular situation.
- (C) select and use an appropriate representation for presenting and displaying relationships among collected data, including line plots, line graphs, stem and leaf plots, circle graphs, bar graphs, box and whisker plots, histograms, and Venn diagrams, with and without the use of technology.

Technology Applications

The student demonstrates knowledge and appropriate use of hardware components, software programs, and their connections. The student is expected to:

- (1)(a) demonstrate knowledge and appropriate use of operating systems, software applications, and communicate and networking components.
- (1)(c) demonstrate the ability to select and use software for a defined task according to quality, appropriateness, effectiveness, and efficiency.
- (1)(f) perform basic software application function including, but not limited to, opening an application program and creating, modifying, printing, and saving documents.

The student uses appropriate computer-based productivity tools to create and modify solutions to problems. The student is expected to:

- (7)(a) plan, create, and edit documents created with a word processor using readable fonts, alignment, page setup, tabs, and ruler settings.
- (7)(b) plan, create, and edit spreadsheet documents using all data types, formulas and functions, and chart information.
- (7)(e) create a document using desktop publishing techniques including, but not limited to, the creation of multi-column or multi-section documents with a variety of text-wrapped frame formats.
- (7)(g) integrate two or more productivity tools into a document including, but not limited to, tables, charts, and graphs, graphics from paint or draw programs, and mail merge.

The student formats digital information for appropriate and effective communication. The student is expected to:

- 10)(a) use productivity tools to create effectiveness document files for defined audiences such as slide shows, poster, multimedia presentations, newsletters, brochures, or reports.
- (11)(a) publish information in a variety of ways including, but not limited to, printed copy, monitor display, Internet documents, and video.

Materials

Advanced Preparation:

- Pre-cut **TEAM CARDS**, enough for one card per group of 2-3 students
- Copies of **Paper Hockey Puck Directions**, **Go Team!**, (optional) **Purple or Orange**, and **Pure Gold** worksheets
- Access to spreadsheet and large monitor or projector/screen for demonstrations
- Access to **GoTeam** and **PureGold** spreadsheet file for each student or pair of students if there is not enough technology available
- Access to a word processor and/or presentation software for each pair of students and **Purple or Orange** file

For each student:

- **Go Team!** and (optional) **Purple or Orange** worksheets
- **PureGold** worksheet

For each student group of 2 -3 students:

- Blank paper or large index cards, one per student
- One **Paper Hockey Puck Directions**, scissors, and one tape measure
- One **TEAM CARD** and one sheet of chart paper, markers

ENGAGE

The Engage portion of the lesson is designed to create student interest in the concepts addressed. This part of the lesson is designed for groups of 2-3 students and then whole group instruction.

1. In small groups, direct the students to take turns flicking a paper hockey puck so that it glides along the surface of the table, measuring the distance traveled (to the nearest inch), and writing that amount with a marker on a blank sheet of paper (or large index card) until each student in the group has a measurement. (Students may design their own paper hockey pucks or use the **Paper Hockey Puck Directions** handout.)

Facilitation Questions

- (Before conducting the activity of flicking the hockey puck) What factors involved in the actual flicking of the hockey puck and measurement of the distance should we standardize for consistency?

Answers may vary. Factors may include how you made the hockey puck, how you hold the hockey puck, starting points, surface on which you are flicking the hockey puck, method of flicking the hockey puck, measuring tools, etc.

Facilitation Questions

- Why should we be concerned with consistency in the way we collect our data? In other words, how might inconsistency affect the validity of our results?
We need to limit the number of variables (factors that may differ) for each event in order to compare distances made when all other factors were the same/controlled. For example, results might be skewed if one group flicked their hockey puck across carpet, and others flick theirs across a slick table.

2. Once each student in the group has recorded his/her measurement, have students line up in numeric order around the room based on the distance traveled by their hockey puck.

Facilitation Questions

- How did you (or how might you) represent the fact that two or more students may have had the same measurement?
Stand behind each other in one spot
- How did you (or how might you) represent the proportional distance between the values of your distances?
Consider themselves as a human number line...proportionally spacing the values.

3. Prompt students to determine the median, mode, and range of their data without technology. Record these statistics on a sheet of chart paper or overhead projector for all to see. Have all students return to their seats.

Facilitation Questions

- How did you (or how might you) determine the median of your data without technology? What impact does the spread of the data have on the value of the median? What is the significance of this value?
Answers may vary. Students could count off from each end of the line to find the middle of their line (median), averaging the values should there be two students in the middle. You might have students raise their hands if their value is above the median, and then do the same for those below the median...to demonstrate that this is a "middle" value. This middle value is not affected by extreme values (outliers) on either end of the data. In this example, it would not be affected by distances that were significantly lower or higher than the rest of the group.

Facilitation Questions

- How did you (or how might you) determine the mode of your data without technology? What is the significance of this value?

Answers may vary. Students could look for where they have students with the same values lined up behind each other (if any). If a set of data has a mode, it indicates that there is a value that occurred multiple times. In this example, this measure of central tendency may not be the best representative of the data unless there is a measurement that occurs many more times than the others.

- How did you (or how might you) determine the range of your data without technology? What is the significance of this value?

Answers may vary. Students could have the students on each end of the line (highest and lowest) find the difference in their measurements. The range helps to describe the spread of the data. In this example, it would tell us if the distances achieved were about the same, or if some students were able to flick for distances that were much longer than others.

4. Input the data into a blank spreadsheet file (using a large monitor or projector for viewing) and demonstrate using formulas to calculate the median, mode, and range.
 - To find median: =median (highlight range of cells with data)
 - To find mode (*If there is more than one mode, Excel returns the smallest mode*): =mode (highlight range of cells with data)
 - To find range: =max (highlight range of cells with data) – min (highlight range of cells with data)
5. Prompt the students to estimate the mean. Verify the estimate using the spreadsheet. To find mean: =average (highlight range of cells with data)
6. Add this statistic to the sheet of chart paper or overhead for all to see.
7. Have students return to their original number line position.
8. Direct the students to form a human histogram.

Facilitation Questions

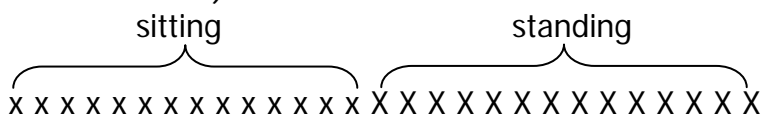
- How might you organize yourselves to create a histogram?
Students in each group should stand behind each other, forming a line ("bar") for each group/range of data.
- What ranges would be appropriate for the bars in our histogram?
Answers may vary based on the data.
- What conjectures can we make based on the spread of the data within this human histogram? In other words, what do we notice about the "shape" of the data?

Answers may vary. Based on the actual spread of the data, students should notice clusters and/or gaps in the spread of the data.

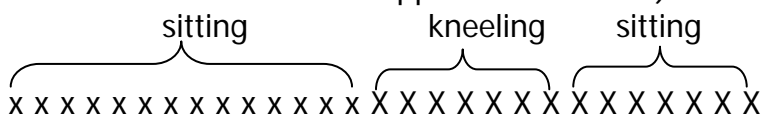
Facilitation Questions

- What kind of information was “lost” when we grouped students together to form the human histogram?
While we know the number of pieces of data in each group, we do not know where in the range for each group the data lies. For example, if there are 5 values within a range of 20 to 30 inches, they could all be closer to 20 inches, 30 inches, or spread throughout the range.

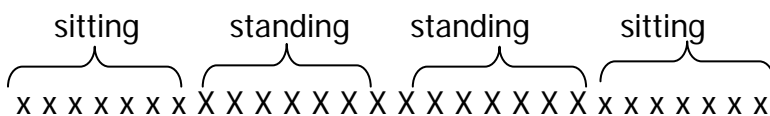
9. Direct the students to form a human box and whisker plot by asking the following questions.
- What was the value of our median? (Identify that person or point between two persons if there is an even number of data points.)
 - Raise your hand if your value is above the median or, in other words, if you are in the upper half of the data? (Once you agree that you have the upper half with their hands raised, ask the lower half to sit on the floor and the upper half can put their hand down.)



- Raise your hand if you are above the median of those standing, in other words, if you are in the upper half of the upper half of the data? (Once you agree that you have the upper half with their hands raised, ask the upper-upper half to sit on the floor and the lower-upper half to kneel.)



- Repeat the procedure to find the upper half of the lower half. Ask the lower - lower half to sit and the upper - lower half to remain standing. Ask the lower - upper half to now stand.



10. Inform the students that they will create histograms and box and whisker plots using technology in the next activity. They will use this “human” version to help understand what is happening “behind the scenes” as the technology creates the graphs.

Facilitation Questions

- How would you describe the data for the group of students still standing? What is the significance of this group?
Answers may vary. Students should note that those standing represent the middle half of the data. Essentially this is the middle group that ignores (does not include) the highs and the lows (outliers).

11. Prompt students to return to their seats.

EXPLORE

The Explore portion of the lesson provides the student with an opportunity to be actively involved in the exploration of the mathematical concepts addressed. This part of the lesson is designed for groups of 2-3 students initially and moving to pairs or individual investigation.

1. Divide the class into groups (teams) of 2-3 students each.
2. Give each group a **TEAM CARD** with a list of numbers that represents the number of hits last season by each person on a particular baseball team. (Duplicate sets of data will be distributed to promote comparing/contrasting comments later in the activity.)
3. Ask each group to calculate the mean (average) of the number of hits their team had last season.
4. Prompt each group to write their team name, list of hits, and the mean on a piece of chart paper.

Facilitation Questions

- What do you notice about each team?
The mean is the same for each team.
- If all three teams had the same mean (average) number of hits last season, what other statistical measures might you examine in order to distinguish between the teams?
Answers may vary. Lead students to classify teams by measurements such as median, mode, range or the spread of the data.

5. Prompt students (or pairs of students if there is not enough technology available) to open the spreadsheet **GoTeam**. Point out that there are four “worksheets” within the file (Blue Team, Green Team, Red Team, Graphs). Have students click on the various tabs toward the bottom of the screen to become familiar with moving between the pages.
6. Distribute the **Go Team** worksheet. (Include the optional page where students may record a summary of their spreadsheet work to each student, if printing capabilities are not available.) If printing student spreadsheets is an option, wait to distribute this worksheet until step 11.
7. In the **Blue Team** worksheet (spreadsheet), ask the students to fill in the number of hits (using the chart paper/Team Card) for each of the 16 players (in any order) in the light yellow cells (B7 – B22).

Facilitation Questions

- How will the order in which you input the data affect the calculations of central tendencies by the technology?
Answers may vary. One benefit of using technology to calculate central tendencies is that order of input does not matter. This would be especially helpful if there were a large number of data points.

8. Once the data has been entered in cells B7-B22, students must sort the data in order for the box and whisker plot to graph correctly. (*Excel uses a different formula to find the lower and upper quartiles than the state will use to assess students. To make the box and whisker plot match the state expectations, certain formulas were entered into the spreadsheet. These formulas will only work if the data is sorted in ascending order.*)
9. Guide the students through inserting the statistical formulas in the light yellow cells in column F. (NOTE: Formula hints will be visible when the cursor is over that cell.)
10. Prompt the students to input the data and formulas for the Green Team and the Red Team, in order for the technology to generate the remaining histograms and box and whisker plots.
11. Distribute the **GoTeam** worksheet with questions 1 – 7 to each student.
12. Prompt the students to use the Graphs worksheet to help them answer the 7 questions on the **Go Team** worksheet.
13. Question #7 will be addressed further in the “Explain” phase.

EXPLAIN

The Explain portion of the lesson is directed by the teacher to allow the students to formalize their understanding of the TEKS addressed in the lesson.

1. Debrief the concept of median using the following questions.

Facilitation Questions

- The median is also referred to as the 50th percentile. Why is this?
50th percentile refers to the value such that, if the data points are sorted from least to greatest, 50% of the data points are less than this value and 50% of the data points are greater than this value.
- Where is the median in your data list? How many values are less than the median and how many are greater than the median?
The median is 50 hits. There are 8 players with fewer than 50 hits, and 8 players with more than 50 hits.

Facilitation Questions

- If the median of the entire group is the 50th percentile, what is the significance of the 25th percentile? How many values are less than the 25th percentile? How many values are greater than the 25th percentile?
This is the median of the lower half of the data. 25% of the data points are less than this value, leaving 75% of the data points to be greater than this value. In this case there are 4 values below the 25th percentile and 12 values above the 25th percentile.
- If the median of the entire group is the 50th percentile, what is the significance of the 75th percentile? How many values are less than the 75th percentile? How many values are greater than the 75th percentile?
This is the median of the upper half of the data. 75% of the data points are less than this value, leaving 25% of the data points to be greater than this value. In this case there are 12 values below the 75th percentile and 4 values above the 75th percentile.
- What is the statistical significance of the median of any group of data? In other words, why does knowing the median of a group of data give us more information than just knowing the average (mean)?
The median represents the "middle" of the data, once the data has been ordered. Since the value of the median is a result of position alone, it is not affected by outliers, whereas even one or two outliers (data points that are significantly higher or lower than the rest of the data) might skew the mean.
- How is this discussion of median versus mean related to the data we have on the number of hits for each player on a team?
The median number of hits is the value that represents the number of hits that is in the "middle" once the values have been ordered. Since this value is based on position alone, it is not as likely to be affected should certain players suddenly have more or less hits.

2. Using a large monitor or projector/screen for viewing, demonstrate how the median is not affected by outliers.

Facilitation Questions

- Since the median and mean numbers of hits for the Blue team are close to the same values (50 and 50.7 respectively), which would change the most if your best hitter was traded for someone with 500 hits? Justify your prediction.
Answers may vary. The median would not change because the data points would not change in position if the largest value is exchanged for an even larger value. The mean would change more because there would be a larger total number of hits, therefore a larger average when those hits were divided out among the players. Type "500" in place of the "120" in the computer and note the lack of change in the median and the significant change in the mean.

Facilitation Questions

- With this exchange of players, does the median (50) or the mean (74.4) better describe the data for the team? Justify your selection.
The median is a better descriptor of the data. Justifications may vary, but students should note that there are very few players on the team that are hitting at or above the mean, while there are still one-half of the players hitting at or above the median (and the other one-half hitting at or below the median).
- How is having the data in the computer helping us justify our thoughts about referring to medians versus means?
Answers may vary. As we are changing values, the computer instantaneously recalculates the central tendencies so that we can focus our discussion on the meaning and significance of each value, without having to get “bogged down” in the recalculations.
NOTE: Ensure that students return the “500” value to “120” before proceeding.

3. Prompt the students to explain how the median is reflected in each graphical representation on the **Graphs** worksheet.

Facilitation Questions

- How is the median represented in each of the histograms on the **Graphs** worksheet?
The median value is contained within the bar representing the group of data on the “41 – 60” bar on each histogram.
- If you did not know the value of the median from the list of data, how could you determine it from the histogram?
By counting frequencies for each bar, you would be able to narrow it down to the appropriate bar, but you would not be able to identify the exact median from the histograms alone.
- How is the median represented in each of the box and whisker plots on the **Graphs** worksheet?
The median value is represented by the line within the box on each box and whisker plots.
- If you did not know the value of the median from the list of data, how could you determine it from the box and whisker plot?
By looking at the position of the line within the box and the number line associated with the plot, you could find the value of the median.

4. Prompt the students to make conjectures about the mean, given the median and the shape of the data in the graphical representations.

Facilitation Questions

- If you had to make an estimate of the mean (average) number of hits, would you rather base your estimate on the histogram or the box and whisker plot? Justify your answer.

Answers may vary. In box and whisker plots, it is easy to estimate the median and then make conjectures about the value of the mean based on the shape of the data around the median. Because of grouping data in ranges on the histogram, it may make the median harder to pinpoint, therefore making it more difficult to determine the relationship between the mean and median.

- Knowing that the mean number of hits for each team is about 50.7, explain why this value “makes sense” based on what you know about the median and the shape of the data in the box and whisker plot.

Possible answers:

Blue team – One might estimate that the value of the mean would be fairly close to the value of the median since the shape of the data in the box and whisker plot indicates that the data on either side of the median is somewhat evenly spread...possibly skewing just a little above the median since there is a little bit larger spread in the data above the median than in the data below the median. You can see this by noting that the right whisker on the box and whisker plot is longer.

Green team- Similar to the discussion about the Blue team, but this time the data below the median is more spread out, leading you to estimate that the mean is a little less than the median.

Red team – While the right whisker is longer than the left, the size of the box to the right of the median indicates that those values are very close to the median. Looking at the spread of the data to the left of the median, one could determine that since these values tend to be farther from the median, the mean would be less than the median.

5. Debrief the concept of range by asking the following Facilitation Questions.

Facilitation Questions

- What is the statistical significance of the range of any group of data? In other words, why does knowing the range of a group of data give us more information than just knowing the average (mean)?

The range is the difference between the maximum and minimum values in a set of data. The range gives you an idea of the spread of the data. The smaller the range, the closer the values of the data points are to each other. As the range increases, so does the spread between the values of the data points.

- How is this discussion of range related to the data we have on the number of hits for each player on a team?

The range in numbers of hits for a team will help us distinguish between teams where the number of hits for the players are more consistent (closer together) and teams where they are not as consistent (some players have significantly more hits than other players on the same team).

6. Prompt the students to make conjectures about the range, given the shape of the data in the graphical representations.

Facilitation Questions

- Would you feel more confident estimating the range of a set of data from a histogram or a box and whisker plot? Justify your answer.

Because some values might get "lost" in the bars at the extremes of the histogram, it is sometimes difficult to closely estimate the range given only this graphical representation...but you can determine a "ballpark" range.

- How is the range reflected in each of the histograms on the **Graphs** worksheet ?

Blue team - While looking at the histogram alone you could only estimate the range to be between 81 and 120, an exact range can be seen on the box and whisker plot by comparing the ends of the whiskers.

Green team - While looking at the histogram alone you could only estimate the range to be between 41 and 80, an exact range can be seen on the box and whisker plot by comparing the ends of the whiskers.

Red team - While looking at the histogram alone you could only estimate the range to be between 81 and 120, an exact range can be seen on the box and whisker plot by comparing the ends of the whiskers.

7. Debrief the concept of mode using the following questions.

Facilitation Questions

- Why does knowing the mode of a group of data give us more information than just knowing the average (mean)?
The mode is the data point with the greatest frequency. If a set of data has a mode, it tells you which value was most common. Depending on the frequency, the value of the mode may or may not impact the mean.
- How is this discussion of mode related to the data we have on the number of hits for each player on a team?
The mode in numbers of hits for a team only tells us if a particular number of hits occurs more often than others. In this case, it would not be uncommon for a set of data to have no mode because there are so many possibilities for numbers of hits.
- If you did not know the value of the mode from the list of data, how could it be determined from the histogram or box and whisker plot alone?
The specific mode is not evident on either representation.

8. Prompt the students to make conjectures about any outliers or clusters, given the shape of the data in the graphical representations.

Facilitation Questions

- For each team, are there any outliers or clusters of data? If so, how are they represented in the histogram and box and whisker plot?
Possible answers:
Blue Team – Two players have a significantly higher number of hits (110 and 120) as seen by the gap in the histogram and the long right whisker on the box and whisker plot. According to the histogram, only one player falls within the range including both the mean and median, leading you to believe the data is spread out.
Green Team – Both the histogram and box and whisker plot show that the vast majority of the players have numbers of hits close to both the mean and median. According to the box and whisker plot, the “middle half” of the data is compressed between the upper 40s and 60, leading you to believe that the numbers of hits for those players are very close together.
Red Team – While the numbers of hits for players are not as tight around the mean and median as the green team, there are still many players clustered within the upper 30s and 60.

9. To answer question #7 on their **Go Team** sheets, prompt the students to use what they know about the central tendencies and range, as well as the graphical representations to explain which team you would like to join.

Facilitation Questions

- Now that we have explored these graphical representations further, think about your response to #7 on your **Go Team** worksheet. If you had the opportunity to join any of these teams for next season, which would it be? Explain.

Answers may vary. Look for explanations that use the data as their justification such as--

The Blue and Red Teams both have players having more than 100 hits.

The Red Team has the highest median.

The Green Team has more consistency between players when it comes to number of hits.

The Blue Team has a number of players hitting well above the median.

ELABORATE

The Elaborate portion of the lesson provides an opportunity for the student to apply the concepts of the TEKS within a new situation. This part of the lesson is designed for pairs of students.

- Pair up the students, to foster student conversation and extend their thoughts as they interpret graphical representations.
- Direct each pair of students to open **Purple or Orange** with their word processor. NOTE: They need the file copy (not just a hard copy) in order to have the ability to cut and paste the graphical representations in their newsletter or slide show. *(Optional) Distribute a hard copy of the **Purple or Orange** file to each student for reference.*
- Prompt students to summarize the directions for the task, giving them additional directions on how to name and save their newsletter or slide show.
- When half of the work time remains, have the student pairs go on a quick (5 - 10 minutes) "spy mission." (For each pair, one student will stay with the work/computer to share their work/thoughts with others, and the other will visit with other students to "compare notes" and possibly hear a different point of view or get a technology tip. The pairs will reunite and continue their work on their newsletter or slide show.)
- After the students have completed the "spy mission," prompt students to defend their answer to #5 to the whole group.

Facilitation Questions

- What evidence is there in the graphical representation(s) to defend your position?
Answers may vary.
- If the data point of 100 hits was added, what change (if any) would you notice in the graphical representation(s)?
Answers may vary.
- If you could call the front office of either team, what question(s) might you ask to help you get a better understanding of the data? In other words, what do you wish you knew that you either don't know or are not sure of based on the graphical representation(s) alone?
Answers may vary.

6. Allow each pair of students a short amount of time to make adjustments to their newsletter or slide show based on information they gained from the whole group discussion.
7. Upon completion of the **Purple or Orange** activity, the teacher should use a rubric to assess student understanding of the concepts addressed in this lesson.

EVALUATE

The Evaluate portion of the lesson provides the student with an opportunity to demonstrate his or her understanding of the TEKS addressed in the lesson.

1. Distribute the **Pure Gold** activity sheet to each student.
2. Clarify the location of and saving procedure for the **PureGold** spreadsheet.
3. Upon completion of the **Pure Gold** activity sheet, the teacher should use a rubric to assess student understanding of the concepts addressed in this lesson.

Answers and Error Analysis for selected response questions

Question Number	TEKS	Correct Answer	Conceptual Error	Conceptual Error	Procedural Error	Procedural Error	Guess
1	8.12(C)	B	A	D	C		
2	8.12(C)	C	A	B			D
3	8.12(A)	D	A	B	C		
4	8.12(C)	A	C	D	B		

Box and Whisker Plot and Histogram Spreadsheet

Team Stats

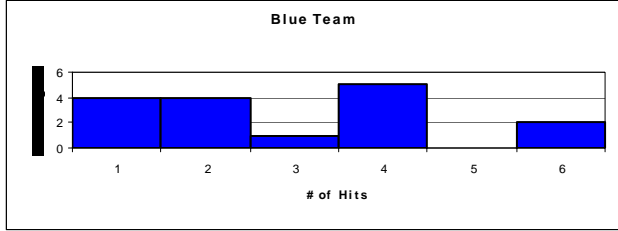
(Give this page to students if printing their spreadsheet is not an option)

See GoTeam-Key spreadsheet.

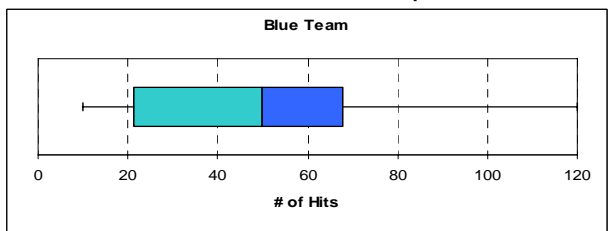
BLUE TEAM

Minimum:	<u>10</u>	Mean:	<u>50.7</u>	25 th %-tile:	<u>19.5</u>
Maximum:	<u>120</u>	Mode:	<u>10</u>	Median:	<u>50</u>
Range:	<u>110</u>			75 th %-tile:	<u>69.5</u>

Sketch the histogram



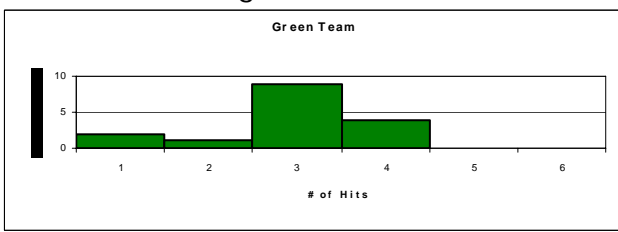
Sketch the box and whisker plot



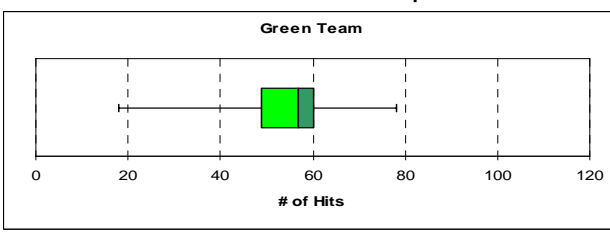
GREEN TEAM

Minimum:	<u>18</u>	Mean:	<u>50.7</u>	25 th %-tile:	<u>48.5</u>
Maximum:	<u>78</u>	Mode:	<u>57, 61</u>	Median:	<u>57</u>
Range:	<u>60</u>			75 th %-tile:	<u>60.5</u>

Sketch the histogram



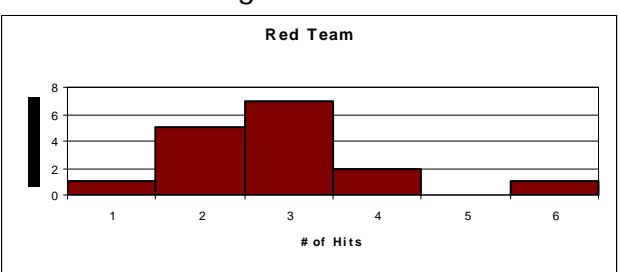
Sketch the box and whisker plot



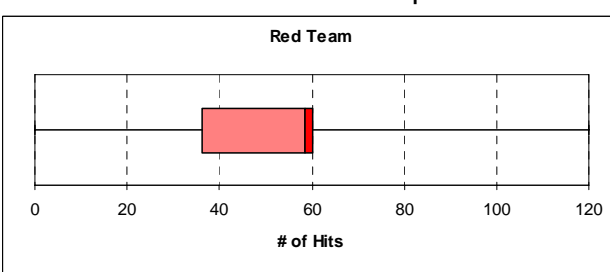
RED TEAM

Minimum:	<u>0</u>	Mean:	<u>50.7</u>	25 th %-tile:	<u>34.5</u>
Maximum:	<u>120</u>	Mode:	<u>60</u>	Median:	<u>58.5</u>
Range:	<u>120</u>			75 th %-tile:	<u>60</u>

Sketch the histogram



Sketch the box and whisker plot



Box and Whisker Plot and Histogram Spreadsheet

Go Team!

Use the terms in the word bank below to complete the statements about the statistics and graphic representations in your spreadsheet. Each term can be used only once.

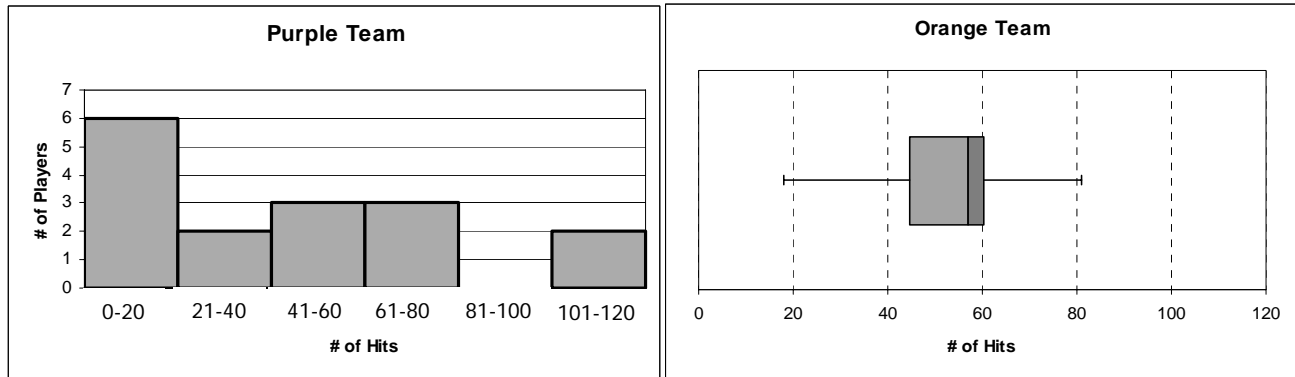
median	Red	mode	Blue
Green	range	outlier	mean

- The box and whisker plot of the **Red** Team has the longest whisker. This is usually an indication that the set of data contains at least one **outlier**.
- The **mode** of the data is the central tendency for which the graphic representations give us the least information.
- The graphic representation with the smallest box (on the box and whisker plot) or with the middle bars significantly taller than the outer bars (on the histogram) for the **Green** Team reflects the fact that the number of hits for many of the players on that team is close to the **median**.
- While the data for each of the three teams is very different, the **mean** number of hits is the same for all.
- The **Blue** and Red Teams both have players with more than 100 hits.
- The **range** of the number of hits was the smallest for the Green Team.

-
-
- If you had the opportunity to join any of these teams for next season, which would it be? Explain using statistics and/or the graphical representation(s) to justify your selection.
Answers may vary.

Purple or Orange? (hard copy of Microsoft Word file)

Below are graphical representations of the number of hits last season by members of the Purple and Orange teams.



Hard Hitting Harold (H^3 for short) has offers to join either the Purple team or the Orange team. H^3 had 100 hits last season.

As a local sports reporter, you have received the task of analyzing the impact for each team, should H^3 join either the Purple or the Orange team. You must base your analysis on what you can gather from the graphical representations you have received.

Use either a word processor to create a newsletter or presentation software to create a slide show that will communicate your interpretations. Copy and paste the graphical representations into your newsletter or slide show and use the drawing tools to help make your points.

Your newsletter or slide show should answer the following questions.

- From the given graphical representations, what do you know about the spread of the data (numbers of hits per player) for the Purple team? for the Orange team? (Include a "discussion" of any clusters, gaps, and/or outliers.)
Answers may vary. Purple...spread between 0 and 120 with half below 41 and a couple of outliers between 101 and 120. Orange...middle half clustered between about 45 and 60...overall spread between about 18 and 81.
- Should H^3 join the team, how would his number of hits (100) impact the current spread of the data for the Purple team? for the Orange team?
Answers may vary. Purple...would appear to fill the 81-100 gap in the histogram, but is just one hit away from being included in the 101-120 group. Orange...would increase the overall spread of the data since the current maximum is around 81...would likely be an outlier, almost 20 greater than the current maximum.

Box and Whisker Plot and Histogram Spreadsheet

3. From the given graphical representations, what do you know about the current range, median, and mean number of hits for the Purple team? for the Orange team?

Answers may vary. Purple...range is somewhere between 80 (20 to 101) and 120 (0 to 120)...median falls between the highest value in the 21-40 bar and the lowest value in the 41-60 bar...mean is likely higher than the median as a result of the outliers over 100. Orange...range is just over 60...median is around 57...mean will likely be less than the median due to the larger spread (not number of data points) of the data below the median

4. Should H³ join the team, what would be the impact on the range, median, and mean number of hits for the Purple team? for the Orange team? *Answers may vary. Purple...will not change the range...will likely increase the median and mean (although hard to say by how much because of the ranges within the bars) because it is greater than both. Orange...will increase the range by around 20,, will likely increase the median and mean because it is greater than both (probably the mean more than the median since the 100 is significantly greater than the current maximum)*

5. In your opinion, which team would benefit the most from having H³ join their team?

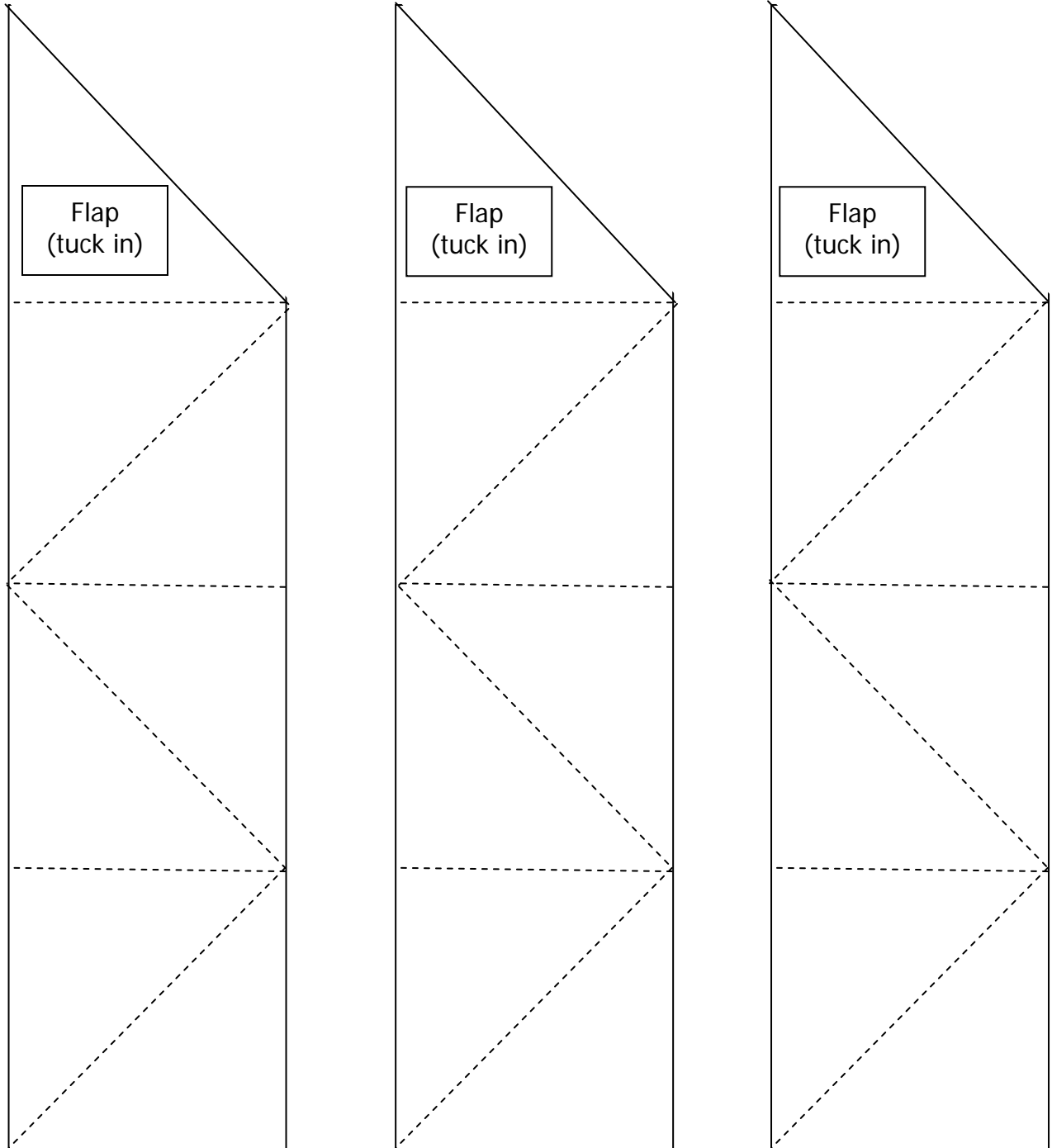
Answers may vary. One could make a case for the Purple team since the 100 would begin to fill in the current gap and/or help to balance against those with very few hits, while the Orange team would gain a player with significantly more hits than the rest of their players, therefore increasing their average number of hits. Accept students' opinions if they can support that opinion with interpretations of the statistics.

6. As an added note or disclaimer, compare and contrast the amount and type of information you were able to get from the histogram versus the box and whisker plot when you addressed questions #1 and 3. What information might you get from a histogram that you would not get from a box and whisker plot? What information might you get from a box and whisker plot that you would not get from a histogram?

Answers may vary. Both help us to see the spread in the data. Since histograms often have more than 4 groupings (6 in this case), gaps and outliers may be more evident. Box and whisker plots, on the other hand, often allow you to estimate the median and range better.

PAPER HOCKEY PUCK DIRECTIONS

Cut out, fold on the dotted lines, and tuck in the flap. (It will form a right triangle when folded.) Each student will need one paper hockey puck.



Use your thumb and pointer of one hand to hold the triangle vertically by the vertices on either side of the hypotenuse. Use your other hand to flick the triangle (hockey puck).

TEAM CARDS

<p>Blue Team</p> <p>10, 120, 15, 16, 23, 27, 66, 39, 40, 60, 10, 61, 64, 73, 77, 110</p>	<p>Blue Team</p> <p>10, 120, 15, 16, 23, 27, 66, 39, 40, 60, 10, 61, 64, 73, 77, 110</p>
<p>Green Team</p> <p>20, 51, 21, 78, 48, 49, 50, 57, 18, 57, 58, 59, 60, 61, 63, 61</p>	<p>Green Team</p> <p>20, 51, 21, 78, 48, 49, 50, 57, 18, 57, 58, 59, 60, 61, 63, 61</p>
<p>Red Team</p> <p>29, 60, 30, 38, 60, 40, 42, 120, 57, 60, 0, 60, 61, 63, 31, 60</p>	<p>Red Team</p> <p>29, 60, 30, 38, 60, 40, 42, 120, 57, 60, 0, 60, 61, 63, 31, 60</p>

Box and Whisker Plot and Histogram Spreadsheet

Team Stats

BLUE TEAM

Minimum: _____ Mean: _____ 25th %-tile: _____
 Maximum: _____ Mode: _____ Median: _____
 Range: _____ 75th %-tile: _____

Sketch the histogram	Sketch the box and whisker plot
----------------------	---------------------------------

GREEN TEAM

Minimum: _____ Mean: _____ 25th %-tile: _____
 Maximum: _____ Mode: _____ Median: _____
 Range: _____ 75th %-tile: _____

Sketch the histogram	Sketch the box and whisker plot
----------------------	---------------------------------

RED TEAM

Minimum: _____ Mean: _____ 25th %-tile: _____
 Maximum: _____ Mode: _____ Median: _____
 Range: _____ 75th %-tile: _____

Sketch the histogram	Sketch the box and whisker plot
----------------------	---------------------------------

Go Team!

Use the terms in the word bank below to complete the statements about the statistics and graphic representations in your spreadsheet. Each term can be used only once.

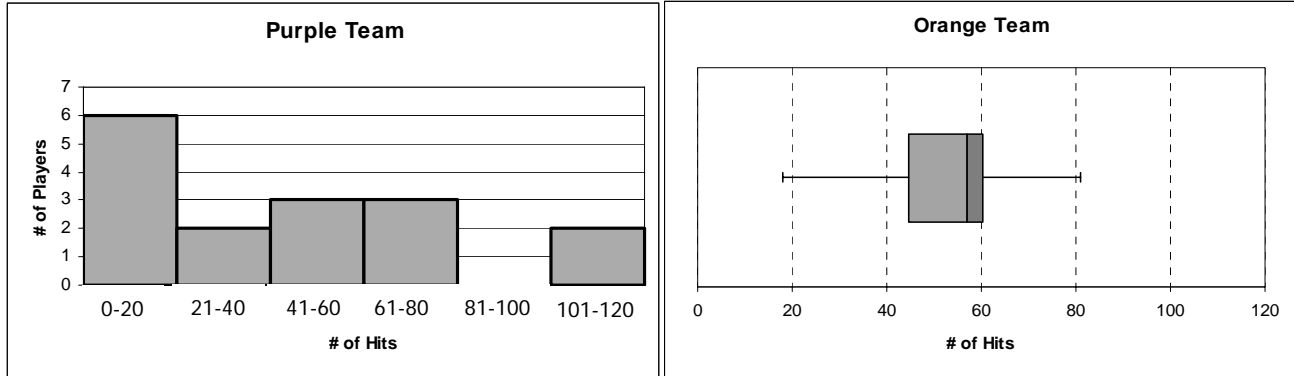
median	Red	mode	Blue
Green	range	outlier	mean

1. The box and whisker plot of the _____ Team has the longest whisker. This is usually an indication that the set of data contains at least one _____ .
2. The _____ of the data is the central tendency for which the graphic representations give us the least information.
3. The graphic representation with the smallest box (on the box and whisker plot) or with the middle bars significantly taller than the outer bars (on the histogram) for the _____ Team reflects the fact that the number of hits for many of the players on that team is close to the _____.
4. While the data for each of the three teams is very different, the _____ number of hits is the same for all.
5. The _____ and Red Teams both have players with more than 100 hits.
6. The _____ of the number of hits was the smallest for the Green Team.

7. If you had the opportunity to join any of these teams for next season, which would it be? Explain using statistics and/or the graphical representation(s) to justify your selection.

Purple or Orange? (hard copy of Microsoft Word file)

Below are graphical representations of the number of hits last season by members of the Purple and Orange teams.



Hard Hitting Harold (H^3 for short) has offers to join both the Purple team and the Orange team. H^3 had 100 hits last season.

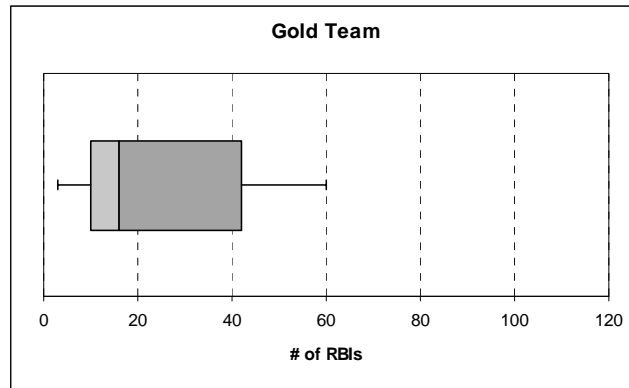
As a local sports reporter, you have received the task of analyzing the impact for each team, should H^3 join either the Purple or the Orange team. You must base your analysis on what you can gather from the graphical representations you have received.

Use either a word processor to create a newsletter or presentation software to create a slide show that will communicate your interpretations. Copy and paste the graphical representations into your newsletter or slide show and use the drawing tools to help make your points. Your newsletter or slide show should answer the following questions.

1. From the given graphical representations, what do you know about the spread of the data (numbers of hits per player) for the Purple team? for the Orange team? (Include a "discussion" of any clusters, gaps, and/or outliers.)
2. Should H^3 join the team, how would his number of hits (100) impact the current spread of the data for the Purple team? for the Orange team?
3. From the given graphical representations, what do you know about the current range, median, and mean number of hits for the Purple team? for the Orange team?
4. Should H^3 join the team, what would be the impact on the range, median, and mean number of hits for the Purple team? for the Orange team?
5. In your opinion, which team would benefit the most from having H^3 join their team?
6. As an added note or disclaimer, compare and contrast the amount and type of information you were able to get from the histogram versus the box and whisker plot when you addressed questions #1 and 3. What information might you get from a histogram that you would not get from a box and whisker plot? What information might you get from a box and whisker plot that you would not get from a histogram?

Pure Gold

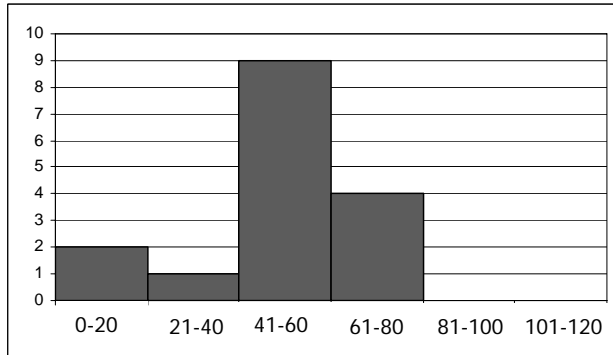
You have just been hired as the manager of the Gold Team. A plot of the number of RBIs (runs batted in) of your team is shown below.



Your first job as team manager is to add 3 players (to replace 3 that retired) to the team. You must meet these goals.

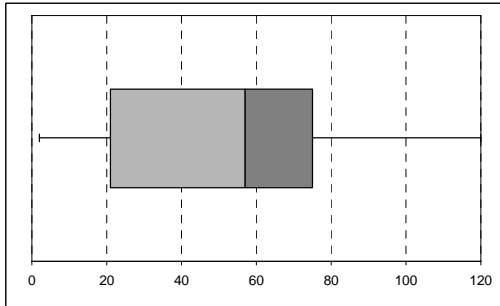
- Do not increase the current range of RBIs.
 - Keep the various numbers of RBIs as clustered around the median as possible.
- a. Open the **PureGold** spreadsheet to see the RBI statistics on your current players and the players that are available to join your team.
 - b. Add 3 players to get the desired results.
 - c. Prepare a statement for the press that lists the RBIs of the players you added and describes the impact of these additions on each of the following statistical measures for your team.
 - d. Justify your statement by including the amount of change (if any) from the original statistics and original box and whisker plot, along with how these additions might benefit the team.
 - Range
 - Median
 - Mean

1.

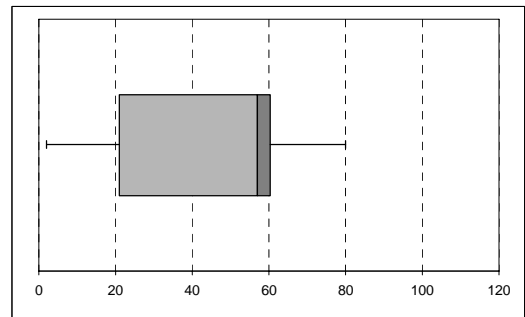


Which of the following box and whisker plots would contain data similar to the histogram above?

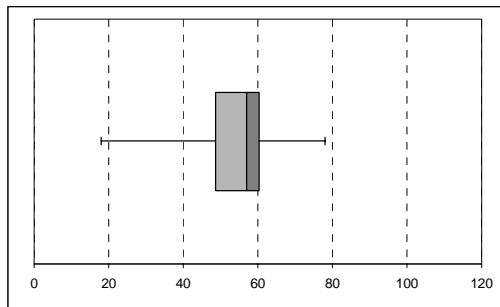
A.



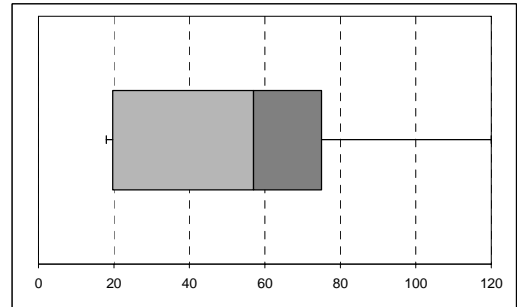
C.



B.



D.

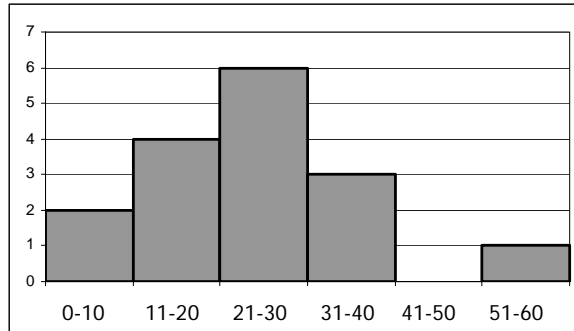


2. Which statistical measure is NOT evident on a box and whisker plot?

- A. range
- B. median
- C. mode
- D. all are evident

Use the information below to answer questions 3 and 4.

A police officer sat on the side of the road and monitored the speed of the traffic with a radar gun. The histogram below represents the speeds of the first sixteen cars to go by.



3. What was the range in speed of the cars?
 - A. 60 mph
 - B. 6 mph
 - C. 50 mph
 - D. cannot be determined from the graph

4. If the road the officer was monitoring was a school zone (speed limit of 20 mph), how many of those cars were speeding?
 - A. 10
 - B. 40
 - C. 6
 - D. not enough information

Mathematics

8.12 The student uses statistical procedures to describe data. The student is expected to:

- (A) select the appropriate measure of central tendency or range to describe a set of data and justify the choice for a particular situation.
- (C) select and use an appropriate representation for presenting and displaying relationships among collected data, including line plots, line graphs, stem and leaf plots, circle graphs, bar graphs, box and whisker plots, histograms, and Venn diagrams, with and without the use of technology.

Technology Applications

The student demonstrates knowledge and appropriate use of hardware components, software programs, and their connections. The student is expected to:

- (1)(a) demonstrate knowledge and appropriate use of operating systems, software applications, and communicate and networking components.
- (1)(c) demonstrate the ability to select and use software for a defined task according to quality, appropriateness, effectiveness, and efficiency.
- (1)(f) perform basic software application function including, but not limited to, opening an application program and creating, modifying, printing, and saving documents.

The student uses appropriate computer-based productivity tools to create and modify solutions to problems. The student is expected to:

- (7)(a) plan, create, and edit documents created with a word processor using readable fonts, alignment, page setup, tabs, and ruler settings.
- (7)(b) plan, create, and edit spreadsheet documents using all data types, formulas and functions, and chart information.
- (7)(e) create a document using desktop publishing techniques including, but not limited to, the creation of multi-column or multi-section documents with a variety of text-wrapped frame formats.
- (7)(g) integrate two or more productivity tools into a document including, but not limited to, tables, charts, and graphs, graphics from paint or draw programs, and mail merge.

The student formats digital information for appropriate and effective communication. The student is expected to:

- 10)(a) use productivity tools to create effectiveness document files for defined audiences such as slide shows, poster, multimedia presentations, newsletters, brochures, or reports.
- (11)(a) publish information in a variety of ways including, but not limited to, printed copy, monitor display, Internet documents, and video.

Materials

Advanced Preparation:

- Access to **TI-73** and large monitor or projector/screen for demonstrations
- Pre-cut **TEAM CARDS**, enough for one card per group of 4-5 students
- Copies of **Paper Hockey Puck Directions**, **Team Stats**, **Go Team!**, **Purple or Orange**, and **Pure Gold** worksheets
- Access to a word processor and/or presentation software for each pair of students – load with **Purple or Orange** file
- Access to a **TI-73** for each student or pair of students
- Transparency of each **TEAM CARD**

For each student:

- **Team Stats**, **Go Team!** and **Purple or Orange** worksheets
- **PureGold** worksheet

For each student group of 4 - 5 students:

- Blank paper or large index cards, one per student
- One **Paper Hockey Puck Directions**, scissors, and one tape measure
- One **TEAM CARD** and one sheet of chart paper, markers

ENGAGE

The Engage portion of the lesson is designed to create student interest in the concepts addressed. This part of the lesson is designed for groups of 4 - 5 students and then whole group instruction.

1. In small groups, direct the students to take turns flicking a paper hockey puck so that it glides along the surface of the table, measuring the distance traveled (to the nearest inch), and writing that amount with a marker on a blank sheet of paper (or large index card) until each in the group has a measurement. (Students may design their own paper hockey pucks or use the **Paper Hockey Puck Directions** handout.)

Facilitation Questions

- (Before conducting the activity of flicking the hockey puck) What factors involved in the actual flicking of the hockey puck and measurement of the distance should we standardize for consistency?
Answers may vary. Factors may include how you made the hockey puck, how you hold the hockey puck, starting points, surface on which you are flicking the hockey puck, method of flicking the hockey puck, measuring tools, etc.
- Why should we be concerned with consistency in the way we collect our data? In other words, how might inconsistency affect the validity of our results?
We need to limit the number of variables (factors that may differ) for each event in order to compare distances made when all other factors were the same/controlled. For example, results might be skewed if one group flicked their hockey puck across carpet, and others flick theirs across a slick table.

2. Have the students take their paper with their measurement and line up in numeric order around the room.

Facilitation Questions

- How did you (or how might you) represent the fact that two or more students may have had the same measurement?
Stand behind each other in one spot
- How did you (or how might you) represent the proportional distance between the values of your distances?
Consider themselves as a human number line...proportionally spacing the values.

3. Prompt the students to determine the median, mode, and range of their data without technology. Record these statistics on a sheet of chart paper or overhead for all to see.
4. Prompt students to return to their seats.
5. Input the data into the **TI-73** List 1 (using a large monitor or projector for viewing) and demonstrate using formulas to calculate median, mean, and mode. You will need to subtract the minimum value from the maximum value to determine the range.

Facilitation Questions

- How did you (or how might you) determine the median of your data without technology? What impact does the spread of the data have on the value of the median? What is the significance of this value?

Answers may vary. Students could count off from each end of the line to find the middle of their line (median), averaging the values should there be two students in the middle. You might have students raise their hand if their value is above the median, and then do the same for those below the median...to further demonstrate that this is a "middle" value. This middle value is not affected by extreme values (outliers) on either end of the data. In this example, it would not be affected by distances that were significantly lower or higher than the rest of the group. (Verify the value with the calculator.)

- How did you (or how might you) determine the mode of your data without technology? What is the significance of this value?

Answers may vary. Students could look for where they have students with the same values lined up behind each other (if any). If a set of data has a mode, it indicates that there is a value that occurred multiple times. In this example, this measure of central tendency may not be the best representative of the data unless there is a measurement that occurs many more times than the others.

- How did you (or how might you) determine the range of your data without technology? What is the significance of this value?

Answers may vary. Students could have the students on each end of the line (highest and lowest) find the difference in their measurements. The range helps to describe the spread of the data. In this example, it would tell us if the distances achieved were about the same, or if some students were able to flick for distances that were much longer than others. (Verify the value with the calculator)

6. Have the students return to the front of the room and line up again in order from least to greatest.

7. Ask students the following questions on creating a histogram.

Facilitation Questions

- How might you organize yourselves to create a histogram (a bar graph based on the ranges agreed to before)?

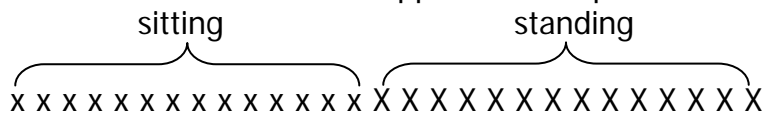
Students in each group should stand behind each other, forming a line ("bar") for each group/range of data.

8. After students have determined a strategy and appropriate ranges for the histogram, direct them to form a human histogram.

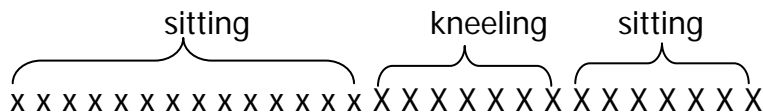
Facilitation Questions

- What conjectures can we make based on the spread of the data within this human histogram? In other words, what do we notice about the “shape” of the data?
Answers may vary. Based on the actual spread of the data, students should notice clusters and/or gaps in the spread of the data.
- What kind of information was “lost” when we grouped students together to form the human histogram?
While we know the number of pieces of data in each group, we do not know where in the range for each group the data lies. For example, if there are 5 values within a range of 20 to 30 inches, they could all be closer to 20 inches, 30 inches, or spread throughout the range.

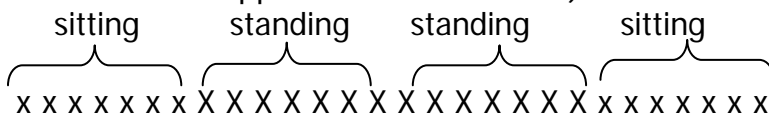
9. Have students return to their number line formation. Lead them in forming a human box and whisker plot.
What was the value of our median? (Identify that person or point between two persons if there is an even number of data points.) Raise your hand if your value is above the median or, in other words, if you are in the upper half of the data? (Once you agree that you have the upper half with their hands raised, ask the lower half to sit on the floor and the upper half can put their hand down.)



Raise your hand if you are above the median of those standing, in other words, if you are in the upper half of the upper half of the data? (Once you agree that you have the upper half of the upper half with their hands raised, ask the upper group to sit on the floor and the lower half to sit on their knees.)



(Repeat the procedure to find the upper half of the lower half. Ask the lower half of the lower half to sit and the upper half of the lower half to remain standing. Ask the lower half of the upper half to now stand.)



Facilitation Questions

- How would you describe the data for the group of students still standing? What is the significance of this group?
Answers may vary. Students should note that those standing represent the middle half of the data. Essentially this is the middle group that ignores (does not include) the highs and the lows (outliers).

10. Students should return to their seats.
11. Sketch the graphical representations on chart paper. Note to the students that they will create histograms and box and whisker plots using technology in the next activity. They will use this “human” version to help understand what is happening “behind the scenes” as the technology creates the graphs.

EXPLORE

The Explore portion of the lesson provides the student with an opportunity to be actively involved in the exploration of the mathematical concepts addressed. This part of the lesson is designed for groups of 4 – 5 students initially, moving to 2 students or individual investigation.

1. Divide the class into 6 groups (teams) of 4 – 5 students each.
2. Give each group a **TEAM CARD** with a list of numbers that represent the number of hits last season by each person on a particular baseball team. Duplicate sets of data will be distributed to promote comparing/contrasting comments later in the activity.
3. Ask each group to calculate the mean (average) of the number of hits their team had last season.
4. Prompt each group to write their team name and list of hits on a piece of chart paper, along with the mean, so that all can see.
5. After noting that all of the teams had the same mean (average) number of hits last season, prompt each group to use something other than the mean to describe the hitting strengths and/or weaknesses that are specific to their team.

Facilitation Questions

- If all three teams had the same mean (average) number of hits last season, what other statistical characteristics might you examine in order to distinguish between the teams?
Answers may vary. Lead students to classify teams by measurements such as median, mode, range or the spread of the data..

6. Distribute a **TI-73** to each student or pair of students.
7. Distribute the **Team Stats** worksheet to each student.

8. Display a transparency of the Blue Team card. Direct the students to fill in the number of hits for each of the 16 players (in any order) in L₁ of their graphing calculators for the Blue Team.

Facilitation Questions

- How will the order in which you input the data affect the calculations of central tendencies?
Answers may vary. One benefit of using a spreadsheet to calculate a central tendency is that the order of input does not matter. This would be especially helpful if there were a large number of data points.

9. Display a transparency of the Green Team and Red Team cards. Direct the students to fill in the number of hits for each of the 16 players (in any order) in L2 for the Green Team and L3 for the Red Team.
10. Guide the students through using the statistical formulas built in the TI-73 to complete the table of stats for the Blue Team.
 - Go to the home screen.
 - Press $\boxed{2\text{nd}}\boxed{\text{LIST}}\boxed{\blacktriangleright}$ to MATH to get the formulas.

```

Ls OPS MATH CALC
1: min(
2: max(
3: mean(
4: median(
5: mode(
6: stdDev(
7: sum(
  
```

- Select the formula and press $\boxed{2\text{nd}}\boxed{\text{LIST}}$ to access the appropriate list number. Press $\boxed{\text{ENTER}}$, then press $\boxed{\text{ENTER}}$ again to see the calculation.

```

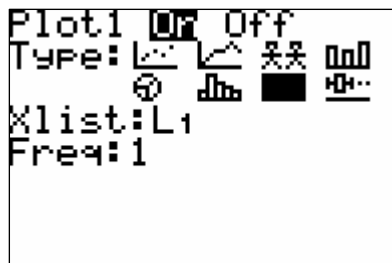
mean(L1  74.4375
█
  
```

- The median is also referred to as the 50th percentile. Why do you think this is so?
50th percentile refers to the value such that, if the data points are sorted from least to greatest, 50% of the data points are less than this value and 50% of the data points are greater than this value. (Students can sort their lists to verify this. Press [2nd][LIST], arrow over to OPS, select SortA, then press [2nd][LIST] to choose the appropriate list. Press [ENTER]. When you return to the list, it should be sorted in ascending order.)
- Where is the median in your data list? How many values are less than the median and how many are greater than the median?
The median is 50 hits. There are 8 players with fewer than 50 hits, and 8 players with more than 50 hits.
- If the median of the entire group is the 50th percentile, what is the significance of the 25th percentile? How many values are less than the 25th percentile? How many values are greater than the 25th percentile?
This is the median of the lower half of the data. 25% of the data points are less than this value, leaving 75% of the data points to be greater than this value. In this case there are 4 values below the 25th percentile and 12 values above the 25th percentile.
- If the median of the entire group is the 50th percentile, what is the significance of the 75th percentile? How many values are less than the 75th percentile? How many values are greater than the 75th percentile?
This is the median of the upper half of the data. 75% of the data points are less than this value, leaving 25% of the data points to be greater than this value. In this case there are 12 values below the 75th percentile and 4 values above the 75th percentile.

11. Direct the students to use their **TI-73** to complete the statistical tables for the Green and Red teams as well.

12. Guide the students through creating a box and whisker plot on the **TI-73**.

- Press [2nd][Y=][ENTER] to turn on Plot1 as a box and whisker plot.



- Press [WINDOW] and adjust the Xmin and Xmax to fit your data (or press ZoomStat).

13. Direct the students to create a box and whisker plot for the Green team on Plot 2 and for the Red team on Plot 3, using their **TI-73**. Students should sketch their box and whisker plots on the **Team Stats** sheet. They will sketch the histograms later in the activity.
14. Prompt the students to use their plots to help them answer the 7 questions on the **Go Team** worksheet.
15. Question #7 will be addressed further in the "Explain" phase.

EXPLAIN

The Explain portion of the lesson is directed by the teacher to allow the students to formalize their understanding of the TEKS addressed in the lesson.

1. Direct the students to sketch a histogram of the data for each team along side the box and whisker plot they sketched on their **Team Stats** worksheet. These graphs will be used to further understand the box and whisker plots. (Suggested interval ranges: 0-20, 21-40, 41-60, 61-80, 81-100, 101-120)
2. Debrief the concept of median using the following questions.

Facilitation Questions

- What is the statistical significance of the median of any group of data? In other words, why does knowing the median of a group of data give us more information than just knowing the average (mean)?
The median represents the "middle" of the data, once the data has been ordered. Since the value of the median is a result of position alone, it is not affected by outliers, whereas even one or two outliers (data points that are significantly higher or lower than the rest of the data) might skew the mean.
- How is this discussion of median versus mean related to the data we have on the number of hits for each player on a team?
The median number of hits is the value that represents the number of hits that is in the "middle" once the values have been ordered. Since this value is based on position alone, it is not as likely to be affected should certain players suddenly have more or less hits.

3. Using a large monitor or projector/screen for viewing, prompt the students to go to the **Blue Team** data in L1 and Plot 1 and demonstrate how the median is not affected by outliers.

Facilitation Questions

- Noting that currently the median and mean numbers of hits for the Blue team are close to the same values (50 and 50.7 respectively), which would change the most if your best hitter was traded for someone with 500 hits? Justify your prediction.

Answers may vary. The median would not change because the data points would not change in position if the largest value is exchanged for an even larger value. The mean would change more because there would be a larger total number of hits, therefore a larger average when those hits were divided out among the players. Students can type "500" in place of the "120" in their spreadsheet and note the lack of change in the median and the significant change in the mean.

- With this exchange of players, does the median (50) or the mean (74.4) better describe the data for the team? Justify your selection.

The median is a better descriptor of the data. Justifications may vary, but students should note that there are very few players on the team that are hitting at or above the mean, while there are still one-half of the players hitting at or above the median (and the other one-half hitting at or below the median).

- How is having the data in a graphing calculator helping us justify our thoughts about referring to medians versus means?

Answers may vary. As we are changing values, we can quickly recalculate the central tendencies so that we can focus our discussion on the meaning and significance of each value, without having to get "bogged down" in the recalculations.

NOTE: Ensure that students return the "500" value to "120" before proceeding.

4. Prompt the students to explain how the median is reflected in each graphical representation on the **GRAPH** screen.

Facilitation Questions

- How is the median reflected in each of the box and whisker plots on the **GRAPH** screen on your calculator?

The median value is represented by the line within the box on each box and whisker plots.

- If you did not already know the value of the median from the list of data, what would you know about it from the box and whisker plot alone?

By looking at the position of the line within the box and the number line associated with the plot, you could find the value of the median.

5. Prompt the students to make conjectures about the mean, given the median and the shape of the data in the graphical representations.

Facilitation Questions

- If you had to make an estimate of the mean (average) number of hits, would you rather base your estimate on the histogram or the box and whisker plot? Justify your answer.
Answers may vary. In box and whisker plots, it is easy to estimate the median and then make conjectures about the value of the mean based on the shape of the data around the median. Because of grouping data in ranges on the histogram, it may make the median harder to pinpoint, therefore making it more difficult to determine the relationship between the mean and median.
- Knowing that the mean number of hits for each team is about 50.7, explain why this value “makes sense” based on what you know about the median and the shape of the data in the box and whisker plot.
Answers may vary.
Blue team – One might estimate that the value of the mean would be fairly close to the value of the median since the shape of the data in box and whisker plot indicates that the data on either side of the median is somewhat evenly spread...possibly skewing just a little above the median since there is a little bit larger spread in the data above the median than in the data below the median. You can see this by noting that the right whisker on the box and whisker plot is longer.
Green team- Similar to the discussion about the Blue team, but this time it is the data below the median that is more spread out, leading you to estimate that the mean is somewhat lower than the median.
Red team – While the right whisker is longer than the left, the size of the box to the right of the median indicates that those values are very close to the median. Looking at the spread of the data to the left of the median, one could determine that since these values tend to be farther from the median, the mean would be less than the median.

6. Debrief the concept of range by asking the following Facilitation Questions.

Facilitation Questions

- What is the statistical significance of the range of any group of data? In other words, why does knowing the range of a group of data give us more information than just knowing the average (mean)?

The range is the difference between the maximum and minimum values in a set of data. The range gives you an idea of the spread of the data. The smaller the range, the closer the values of the data points are to each other. As the range increases, so does the spread between the values of the data points.

- How is this discussion of range related to the data we have on the number of hits for each player on a team?

The range in numbers of hits for a team will help us distinguish between teams where the number of hits for the players are more consistent (closer together) and teams where they are not as consistent (some players have significantly more hits than other players on the same team).

- Prompt the students to make conjectures about the range, given the shape of the data in the graphical representations, using the following questions.

Facilitation Questions

- Would you feel more confident estimating the range of a set of data from a histogram or a box and whisker plot? Justify your answer.

Because some values might get "lost" in the bars at the extremes of the histogram, it is sometimes difficult to closely estimate the range given only this graphical representation...but you can determine a "ballpark" range.

- How is the range reflected in each of the histograms on the **Graphs** page of your spreadsheet?

Blue team - While looking at the histogram alone you could only estimate the range to be between 81 and 120, an exact range can be seen on the box and whisker plot by comparing the ends of the whiskers.

Green team - While looking at the histogram alone you could only estimate the range to be between 41 and 80, an exact range can be seen on the box and whisker plot by comparing the ends of the whiskers.

Red team - While looking at the histogram alone you could only estimate the range to be between 81 and 120, an exact range can be seen on the box and whisker plot by comparing the ends of the whiskers.

8. Debrief the concept of mode using the following questions.

Facilitation Questions

- What is the statistical significance of the mode of any group of data? In other words, why does knowing the mode of a group of data give us more information than just knowing the average (mean)?

The mode is the data point with the greatest frequency. If a set of data has a mode, it tells you which value was most common. Depending on the frequency, the value of the mode may or may not impact the mean.

- How is this discussion of mode related to the data we have on the number of hits for each player on a team?

The mode in numbers of hits for a team only tells us if a particular number of hits occur more often than others. In this case, it would not be uncommon for a set of data to have no mode because there are so many possibilities for numbers of hits.

9. Prompt the students to make conjectures about the range, given the shape of the data in the graphical representations.

Facilitation Questions

- If you did not already know the value of the range from the list of data, what would you know about it from the histogram or box and whisker plot alone?

The specific mode is not evident on either representation.

10. Prompt the students to make conjectures about any outliers or clusters, given the shape of the data in the graphical representations.

Facilitation Questions

- For each team, are there any outliers or clusters of data? If so, how are they reflected in the histogram and box and whisker plot?

Answers may vary.

Blue Team – Two players have a significantly higher number of hits (110 and 120) as see by the gap in the histogram and the long right whisker on the box and whisker plot. According to the histogram, only one player falls within the range including both the mean and median, leading you to believe the data is spread out.

Green Team – Both the histogram and box and whisker plot show that the vast majority of the players have numbers of hits close to both the mean and median. According to the box and whisker plot, the “middle half” of the data is compressed between the upper 40’s and 60, leading you to believe that the numbers of hits for those players are very close together.

Red Team – While the numbers of hits for players is not as tight around the mean and median as the green team, there are still many players clustered within the upper 30s and 60.

11. To answer question #7 on their **Go Team** sheets, prompt the students to use what they know about the central tendencies and range, as well as the graphical representations to explain which team you would like to join.

Facilitation Questions

- Now that we have explored these graphical representations further, think again about your response to #7 on your **Go Team** worksheet. If you got the choice to join any of these teams for next season, which would it be? Explain. (Go to the **Graphs** worksheet to make it easier to see all at the same time.)
Answers may vary. Look for explanations that use the data as their justification such as--
The Blue and Red Teams both have players having more than 100 hits.
The Red Team has the highest median.
The Green Team has more consistency between players when it comes to number of hits.
The Blue Team has a number of players hitting well above the median.

ELABORATE

The Elaborate portion of the lesson provides an opportunity for the student to apply the concepts of the TEKS within a new situation. This part of the lesson is designed for pairs of students.

1. Pair up the students, to foster student conversation and extend their thoughts as they interpret graphical representations.
2. Direct each pair of students to open **Purple or Orange** with their word processor. NOTE: They need the file copy (not just a hard copy) in order to have the ability to cut and paste the graphical representations in their newsletter or slide show.
3. Distribute a hard copy of the **Purple or Orange** file to each student for reference.
4. Prompt students to summarize the directions for the task, giving them additional directions on how to name and save their newsletter or slide show.
5. With about one-half of the work time remaining, have the pairs go on a quick (5 - 10 minutes) "spy mission." For each pair, one student will stay with their work/computer to share their work/thoughts with others, and the other will visit with other students to "compare notes" and possibly hear a different point of view or get a technology tip. The pairs will reunite and continue their work on their newsletter or slide show.
6. After the students are near completion of their newsletter or slide show, prompt students to defend their answer to #5 to the whole group.

Facilitation Questions

- What evidence is there in the graphical representation(s) to defend your position?
Answers may vary.
- If the data point of 100 hits was added, what change (if any) would you notice in the graphical representation(s)?
Answers may vary.
- If you could call the front office of either team, what question(s) might you ask them to help you get a better understanding of the data? In other words, what do you wish you knew that you either don't know or are not sure of based on the graphical representation(s) alone?
Answers may vary.

7. Allow the pairs of students a short amount of time to make adjustments to their newsletter or slide show based on what they gained from the whole group discussion.
8. Upon completion of the **Purple or Orange** activity, the teacher should use a rubric to assess student understanding of the concepts addressed in this lesson.

EVALUATE

The Evaluate portion of the lesson provides the student with an opportunity to demonstrate his or her understanding of the TEKS addressed in the lesson.

1. Distribute **Pure Gold** activity sheet to each student.
2. Upon completion of the **Pure Gold** activity sheet, the teacher should use a rubric to assess student understanding of the concepts addressed in this lesson.

Answers and Error Analysis for selected response questions

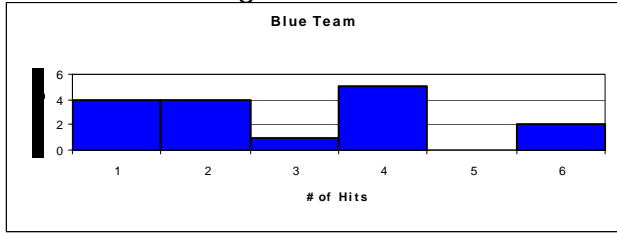
Question Number	TEKS	Correct Answer	Conceptual Error	Conceptual Error	Procedural Error	Procedural Error	Guess
1	8.12(C)	B	A	D	C		
2	8.12(C)	C	A	B			D
3	8.12(A)	D	A	B	C		
4	8.12(C)	A	C	D	B		

Team Stats

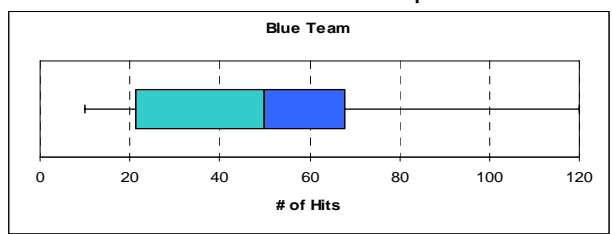
BLUE TEAM

Minimum: 10 Mean: 50.7 25th %-tile: 19.5
 Maximum: 120 Mode: 10 Median: 50
 Range: 110 75th %-tile: 69.5

Sketch the histogram



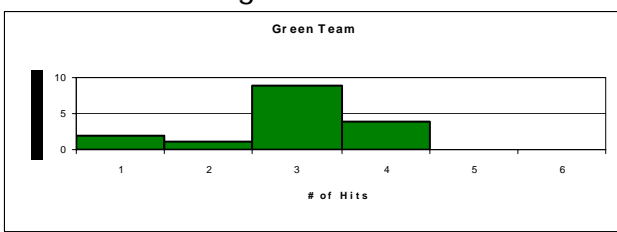
Sketch the box and whisker plot



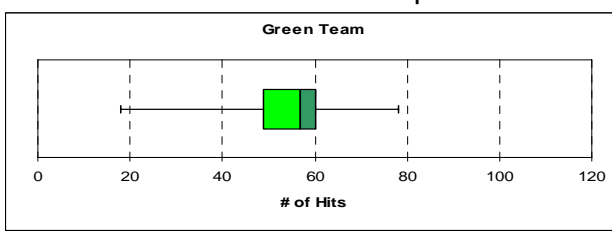
GREEN TEAM

Minimum: 18 Mean: 50.7 25th %-tile: 48.5
 Maximum: 78 Mode: 57, 61 Median: 57
 Range: 60 75th %-tile: 60.5

Sketch the histogram



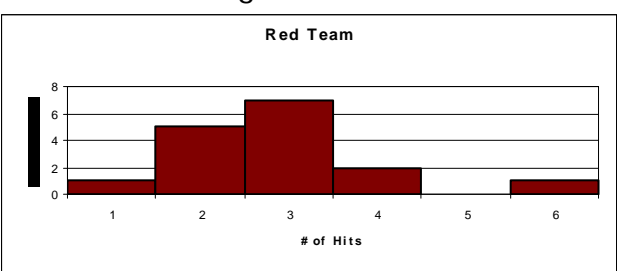
Sketch the box and whisker plot



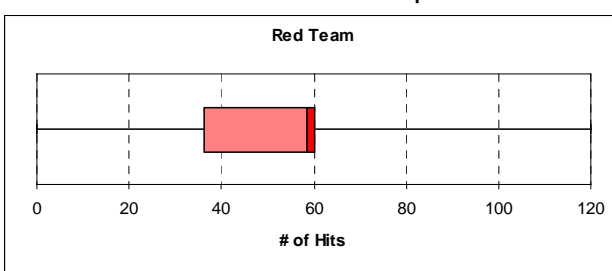
RED TEAM

Minimum: 0 Mean: 50.7 25th %-tile: 34.5
 Maximum: 120 Mode: 60 Median: 58.5
 Range: 120 75th %-tile: 60

Sketch the histogram



Sketch the box and whisker plot



EXPLORE

Go Team!

Use the terms in the word bank below to complete the statements about the statistics and graphic representations in your calculator. Each term can be used only once.

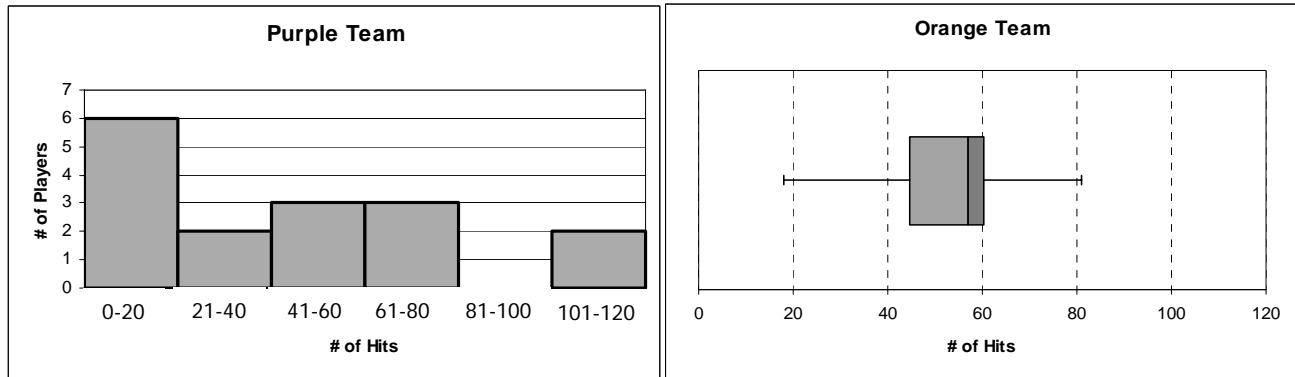
median	Red	mode	Blue
Green	range	outlier	mean

1. The box and whisker plot of the **Red** Team has the longest whisker. This is usually an indication that the set of data contains at least one **outlier**.
2. The **mode** of the data is the central tendency for which the graphic representations give us the least information.
3. The graphic representation with the smallest box (on the box and whisker plot) for the **Green** Team reflect the fact that the number of hits for many of the players on that team is close to the **median**.
4. While the data for each of the three teams is very different, the **mean** number of hits was the same for all.
5. The **Blue** and Red Teams both have players with more than 100 hits.
6. The **range** of the number of hits was the smallest for the Green Team.

7. If you got the choice to join any of these teams for next season, which would it be? Explain using statistics and/or the graphical representation(s) to justify your selection. (Go to the **Team Stats** worksheet to make it easier to see all at the same time.)
Answers may vary.

Purple or Orange? (hard copy of Microsoft Word file)

Below are graphical representations of the number of hits last season by members of the Purple and Orange teams.



Hard Hitting Harold (H^3 for short) has offers to join either the Purple team or the Orange team. H^3 had 100 hits last season.

As a local sports reporter, you have been given the task of analyzing the impact for each team, should H^3 join either the Purple or the Orange team. You must base your analysis on what you can gather from the graphical representations you have been given.

Use either a word processor to create a newsletter or presentation software to create a slide show that will communicate your interpretations. Copy and paste the graphical representations into your newsletter or slide show and use the drawing tools to help make your points.

Your newsletter or slide show should answer the following questions.

- From the given graphical representations, what do you know about the spread of the data (numbers of hits per player) for the Purple team? for the Orange team? (Include a "discussion" of any clusters, gaps, and/or outliers.)
Answers may vary. Purple...spread between 0 and 120 with half below 41 and a couple of outliers between 101 and 120. Orange...middle half clustered between about 45 and 60...overall spread between about 18 and 81.
- Should H^3 join the team, how would his number of hits (100) impact the current spread of the data for the Purple team? for the Orange team?
Answers may vary. Purple...would appear to fill the 81-100 gap in the histogram, but is just one hit away from being included in the 101-120 group. Orange...would increase the overall spread of the data since the current maximum is around 81...would likely be an outlier, almost 20 greater than the current maximum.

Purple or Orange? (hard copy of Microsoft Word file) - continued

3. From the given graphical representations, what do you know about the current range, median, and mean number of hits for the Purple team? for the Orange team?

Answers may vary. Purple...range is somewhere between 80 (20 to 101) and 120 (0 to 120)...median falls between the highest value in the 21-40 bar and the lowest value in the 41-60 bar...mean is likely higher than the median as a result of the outliers over 100. Orange...range is just over 60...median is around 57...mean will likely be less than the median due to the larger spread (not number of data points) of the data below the median

4. Should H³ join the team, what would be the impact on the range, median, and mean number of hits for the Purple team? for the Orange team? *Answers may vary. Purple...will not change the range...will likely increase the median and mean (although hard to say by how much because of the ranges within the bars) because it is greater than both. Orange...will increase the range by around 20...will likely increase the median and mean because it is greater than both (probably the mean more than the median since the 100 is significantly greater than the current maximum)*

5. In your opinion, which team would benefit the most from having H³ join their team?

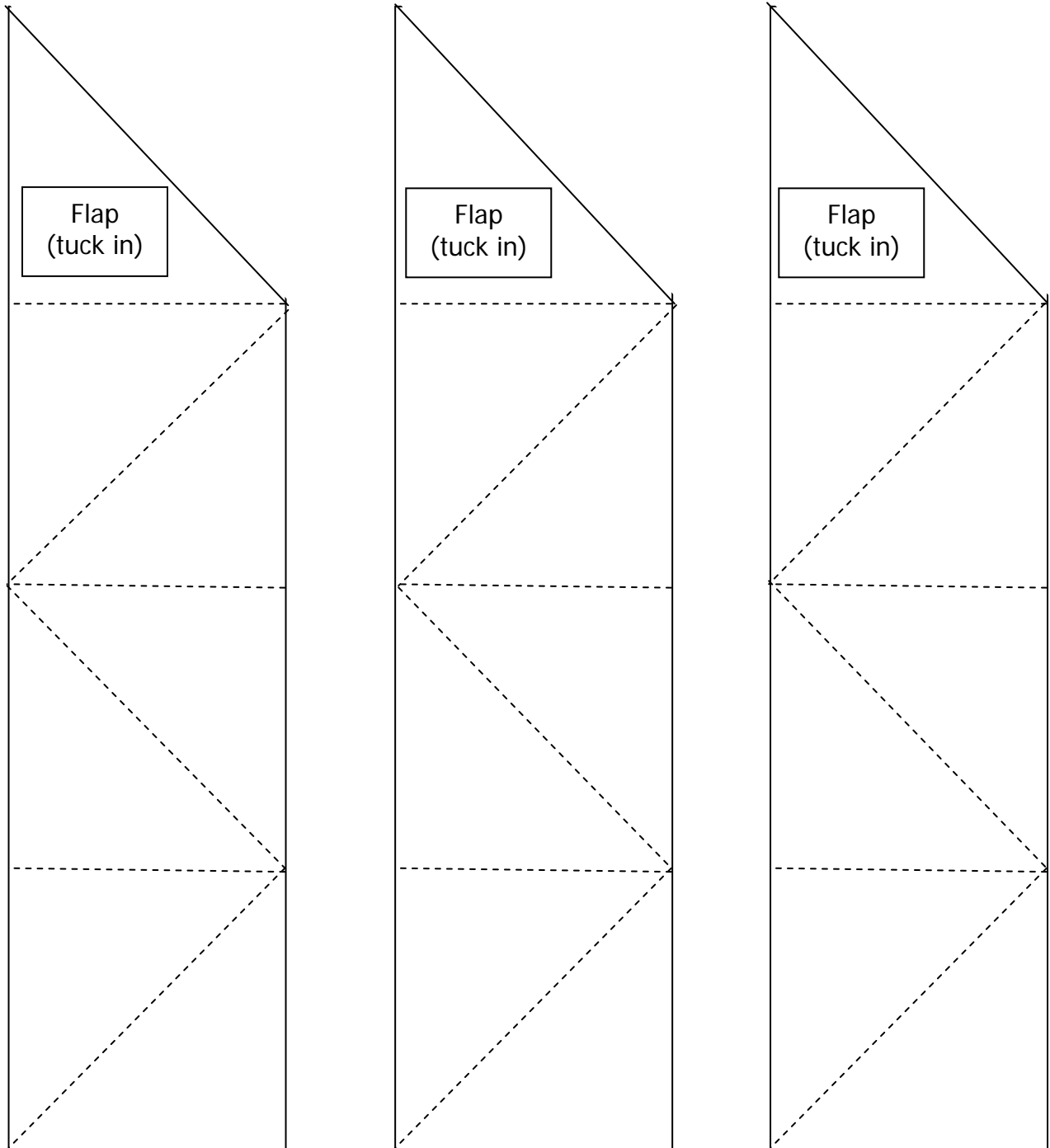
Answers may vary. One could make a case for the Purple team since the 100 would begin to fill in the current gap and/or help to balance against those with very few hits, while the Orange team would gain a player with significantly more hits than the rest of their players, therefore increasing their average number of hits. Opinions should be accepted if the student is able to verbalize how they can back up that opinion with interpretations of the statistics.

6. As an added note or disclaimer, compare and contrast the amount and type of information you were able to get from the histogram versus the box and whisker plot when you addressed questions #1 and 3. What information might you get from a histogram that you would not get from a box and whisker plot? What information might you get from a box and whisker plot that you would not get from a histogram?

Answers may vary. Both help us to see the spread in the data. Since histograms often have more than 4 groupings (6 in this case), gaps and outliers may be more evident. Box and whisker plots, on the other hand, often allow you to better estimate the median and range.

PAPER HOCKEY PUCK DIRECTIONS

Cut out, fold on the dotted lines, and tuck in the flap. (It will form a right triangle when folded.) Each student will need one paper hockey puck.



Use your thumb and pointer of one hand to hold the triangle vertically by the vertices on either side of the hypotenuse. Use your other hand to flick the triangle (hockey puck).

TEAM CARDS

<p>Blue Team</p> <p>10, 120, 15, 16, 23, 27, 66, 39, 40, 60, 10, 61, 64, 73, 77, 110</p>	<p>Blue Team</p> <p>10, 120, 15, 16, 23, 27, 66, 39, 40, 60, 10, 61, 64, 73, 77, 110</p>
<p>Green Team</p> <p>20, 51, 21, 78, 48, 49, 50, 57, 18, 57, 58, 59, 60, 61, 63, 61</p>	<p>Green Team</p> <p>20, 51, 21, 78, 48, 49, 50, 57, 18, 57, 58, 59, 60, 61, 63, 61</p>
<p>Red Team</p> <p>29, 60, 30, 38, 60, 40, 42, 120, 57, 60, 0, 60, 61, 63, 31, 60</p>	<p>Red Team</p> <p>29, 60, 30, 38, 60, 40, 42, 120, 57, 60, 0, 60, 61, 63, 31, 60</p>

Team Stats

BLUE TEAM

Minimum: _____ Mean: _____ 25th %-tile: _____
 Maximum: _____ Mode: _____ Median: _____
 Range: _____ 75th %-tile: _____

Sketch the histogram	Sketch the box and whisker plot
----------------------	---------------------------------

GREEN TEAM

Minimum: _____ Mean: _____ 25th %-tile: _____
 Maximum: _____ Mode: _____ Median: _____
 Range: _____ 75th %-tile: _____

Sketch the histogram	Sketch the box and whisker plot
----------------------	---------------------------------

RED TEAM

Minimum: _____ Mean: _____ 25th %-tile: _____
 Maximum: _____ Mode: _____ Median: _____
 Range: _____ 75th %-tile: _____

Sketch the histogram	Sketch the box and whisker plot
----------------------	---------------------------------

Go Team!

Use the terms in the word bank below to complete the statements about the statistics and graphic representations in your calculator. Each term can be used only once.

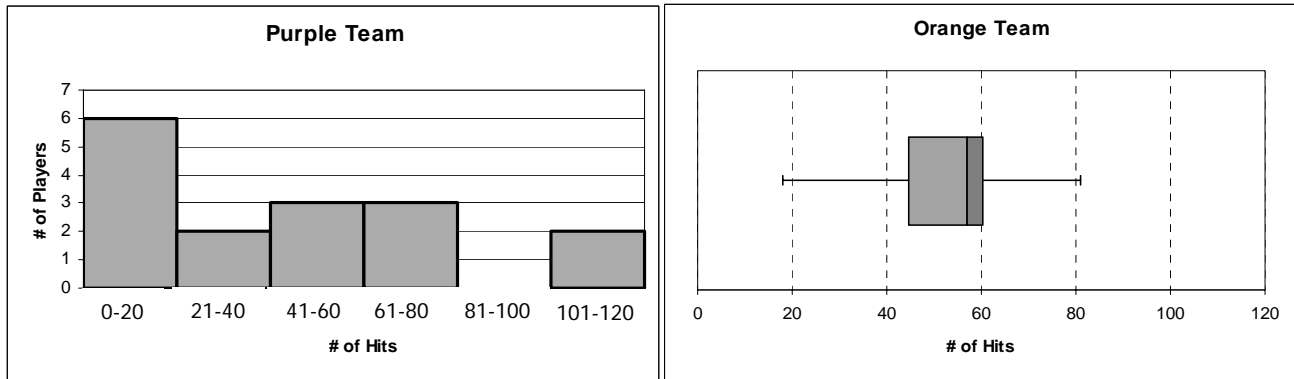
median	Red	mode	Blue
Green	range	outlier	mean

1. The box and whisker plot of the _____ Team has the longest whisker. This is usually an indication that the set of data contains at least one _____ .
2. The _____ of the data is the central tendency for which the graphic representations give us the least information.
3. The graphic representation with the smallest box (on the box and whisker plot) for the _____ Team reflect the fact that the number of hits for many of the players on that team is close to the _____.
4. While the data for each of the three teams is very different, the _____ number of hits was the same for all.
5. The _____ and Red Teams both have players with more than 100 hits.
6. The _____ of the number of hits was the smallest for the Green Team.

7. If you got the choice to join any of these teams for next season, which would it be? Explain using statistics and/or the graphical representation(s) to justify your selection. (Go to the **Graphs** worksheet to make it easier to see all at the same time.)

Purple or Orange? (hard copy of Microsoft Word file)

Below are graphical representations of the number of hits last season by members of the Purple and Orange teams.



Hard Hitting Harold (H^3 for short) has offers to join either the Purple team or the Orange team. H^3 had 100 hits last season.

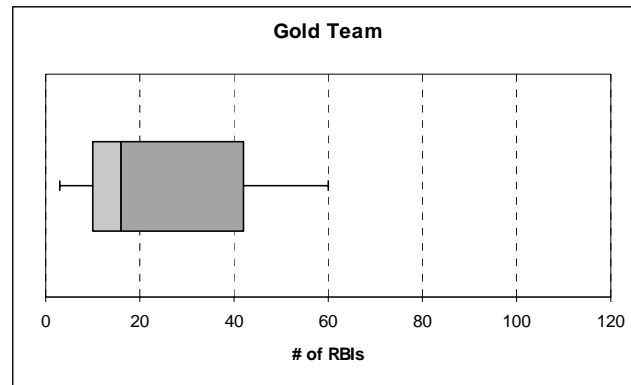
As a local sports reporter, you have been given the task of analyzing the impact for each team, should H^3 join either the Purple or the Orange team. You must base your analysis on what you can gather from the graphical representations you have been given.

Use either a word processor to create a newsletter or presentation software to create a slide show that will communicate your interpretations. Copy and paste the graphical representations into your newsletter or slide show and use the drawing tools to help make your points. Your newsletter or slide show should answer the following questions.

1. From the given graphical representations, what do you know about the spread of the data (numbers of hits per player) for the Purple team? for the Orange team? (Include a "discussion" of any clusters, gaps, and/or outliers.)
2. Should H^3 join the team, how would his number of hits (100) impact the current spread of the data for the Purple team? for the Orange team?
3. From the given graphical representations, what do you know about the current range, median, and mean number of hits for the Purple team? for the Orange team?
4. Should H^3 join the team, what would be the impact on the range, median, and mean number of hits for the Purple team? for the Orange team?
5. In your opinion, which team would benefit the most from having H^3 join their team?
6. As an added note or disclaimer, compare and contrast the amount and type of information you were able to get from the histogram versus the box and whisker plot when you addressed questions #1 and 3. What information might you get from a histogram that you would not get from a box and whisker plot? What information might you get from a box and whisker plot that you would not get from a histogram?

Pure Gold

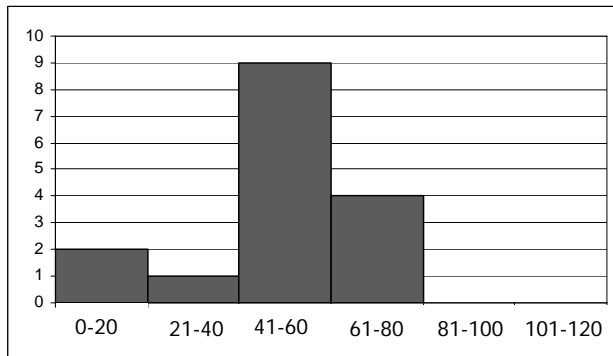
You have just been hired to take over as manager of the Gold Team. Currently, a plot of the number of RBIs (runs batted in) by the members of your team is shown below.



Your first task will be to add 3 players (to replace 3 that retired) to the team to meet these goals.

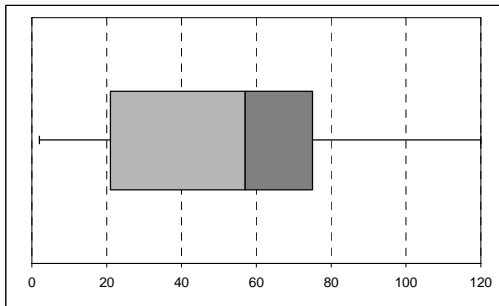
- Do not increase the current range of RBIs.
 - Keep the various numbers of RBIs as clustered around the median as possible.
- a. Input the RBIs for the players currently on the team in L1.
3, 15, 8, 20, 45, 16, 39, 10, 10, 42, 60, 55, 5
 - b. Add 3 players to get the desired results. The RBIs for the players available to join your team are 65, 18, 22, 6, and 30.
 - c. Prepare a statement for the press that lists the RBIs of the players you added and describes the impact of these additions on each of the following statistical measures for your team. Justify your statement by including the amount of change (if any) from the original statistics and original box and whisker plot, along with how these additions might benefit the team.
 - Range
 - Median
 - Mean

1.

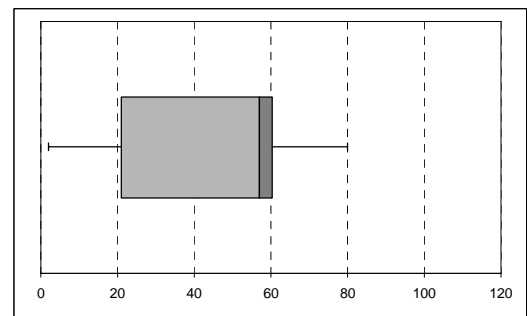


Which of the following box and whisker plots would contain data similar to the histogram above?

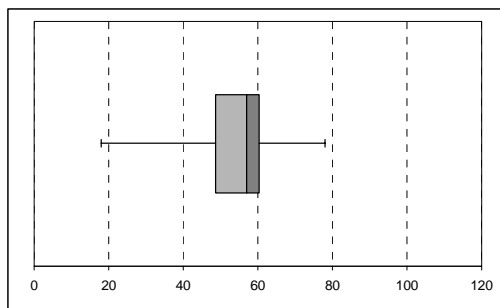
A.



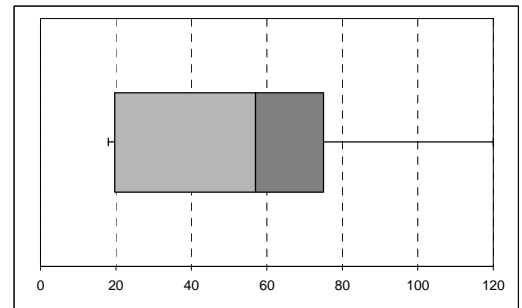
C.



B.



D.

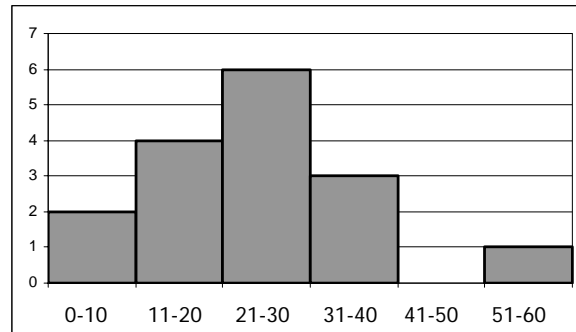


2. Which statistical measure is **not** evident on a box and whisker plot?

- A. range
- B. median
- C. mode
- D. all are evident

Use the following graph to answer questions 3 and 4.

A police officer sat on the side of the road and monitored the speed of the traffic with a radar gun. The histogram below represents the speeds of the first sixteen cars to go by.



3. What was the range in speed of the cars?
 - A. 60 mph
 - B. 6 mph
 - C. 50 mph
 - D. not enough information

4. If the road the officer was monitoring was a school zone (speed limit of 20 mph), how many of those cars were speeding?
 - A. 10
 - B. 40
 - C. 6
 - D. not enough information

Mathematics

8.12 The student uses statistical procedures to describe data. The student is expected to:

- (A) select the appropriate measure of central tendency or range to describe a set of data and justify the choice for a particular situation.
- (B) draw conclusions and make predictions by analyzing trends in scatterplots.
- (C) Select and use an appropriate representation for presenting and displaying relationships among collected data, including line plots, line graphs, stem and leaf plots, circle graphs, bar graphs, box and whisker plots, histograms, and Venn diagrams, with and without the use of technology.

Technology Applications

The student demonstrates knowledge and appropriate use of hardware components, software programs, and their connections. The student is expected to:

- (1)(a) demonstrate knowledge and appropriate use of operating systems, software applications, and communicate and networking components.
- (1)(c) demonstrate the ability to select and use software for a defined task according to quality, appropriateness, effectiveness, and efficiency.
- (1)(f) perform basic software application function including, but not limited to, opening an application program and creating, modifying, printing, and saving documents.

The students uses appropriate computer-based productivity tools to create and modify solutions to problems. The student is expected to:

- (7)(a) plan, create, and edit documents created with a word processor using readable fonts, alignment, page setup, tabs, and ruler settings.

Materials

Advanced Preparation:

- Student copies of **RoundandRound**, **BabyName**, and **WhatName** spreadsheets or copies available on a network
- Internet access to <http://www.ssa.gov/OACT/babynames/>
- Printer access
- Copies of **Round and Round** and **What's In A Name** worksheets for each student

For whole class demonstration:

- Transparencies 1 - 5 (2 copies of Transparency 4)
- Several hula hoops (ideally, one for every 2-3 students)

For each student:

- **Round and Round** worksheet
- Access to **RoundandRound** spreadsheet
- Access to **BabyName** spreadsheet
- **What's In A Name** worksheet
- Access to **WhatName** spreadsheet

ENGAGE

The Engage portion of the lesson is designed to create student interest in the concepts addressed. This part of the lesson is designed for whole class discussion/demonstration.

1. Prompt a student to read **Transparency 1** and ask students to individually consider their prediction.
2. Prompt students to share their predictions with a neighbor before getting responses from the large group.

Facilitation Questions

- What data are you collecting and comparing for each student?
We are looking for, and comparing, the number of revolutions on the first attempt and the number of revolutions on the second attempt.
- Once several students have recorded their predictions on the numbers of revolutions on the first attempt, what will you need to consider when making predictions about the second attempt?
*Answers may vary, but you are looking for the fact that students recognize that, chances are, the change in the number of revolutions between attempts will not be exactly the same for all students—some will increase, some will not. Instead, they should look for trends that describe the change, such as “The number of revolutions for the second attempt **tends** to be _____.”*
- Will your prediction be true for every student? Why or why not?
No. Point out that this exemplifies how “real” data often does not fit into “clean” or exact patterns such as linear patterns. Instead, we have to look for any trends in the data.
- How many pieces of data would you need to make a prediction about the number of revolutions on a second attempt?
Answers may vary, but you are looking for students to recognize that the more data you have, the better defined any correlation will be.

3. Show the table on **Transparency 2** that will be used to collect the data. Ask for 2-3 volunteers to demonstrate what will be recorded in the table. You may want to agree ahead of time as to whether or not each student will be allowed a practice attempt. (If hula-hoops are not available, you can substitute another event such as

paddle-ball, trash can basketball, etc. The goal is to choose an event where the trend in the data is not obvious or that you could make a case for multiple trends. For example, one might make the case that students would have more revolutions with the hula-hoop on their 2nd attempt because of the practice they got during the 1st attempt or that they would have fewer on their 2nd attempt because they were tired from the 1st attempt or that there would be no clear correlation.)

- Pair the students in the class to collect the data for Transparency 2. Each student will take turns being the hula-hooper and the recorder. Record the results on Transparency 2.

Facilitation Questions

- If a student is able to complete 6 revolutions on the first attempt, what could happen on the second attempt? Why?
The number of revolutions could be more than 6, less than 6, or the same as 6 on the second attempt.
- If that same student were able to make 10 revolutions on the second attempt (an increase of 4 over the first attempt), would this mean that the same will be true for the next student? Why?
Possibly, but possibly not—we do not have enough data to make that prediction yet.

EXPLORE

The Explore portion of the lesson provides the student with an opportunity to be actively involved in the exploration of the mathematical concepts addressed. This part of the lesson is designed for groups of 2 students or individual investigation.

- Prompt students (or pairs of students if there is not enough technology available) to open the spreadsheet **RoundandRound**.
- Distribute the **Round and Round** worksheet. Have students follow the directions on the spreadsheet.

Facilitation Questions

- How would you describe any trends that you might see in the scatterplot drawn from the data from your class?
Answers may vary depending on the data collected. Look for statements such as "As the number of revolutions during the 1st attempt increases, the number of revolutions during the 2nd attempt (increases/decreases)." Or "Students tend to _____."

- For part B, inform students these graphs were drawn based on fictitious data, and not their own, as they answer questions #1 – 7 on the worksheet.

Facilitation Questions

- Describe a possible scenario that would produce each of the three scatterplots.

Answers may vary...

Scatterplot A – Most students performed about the same on their first attempt as on their second attempt.

Scatterplot B – Most students did considerably better on their second attempt than on their first attempt.

Scatterplot C – Some students followed the explanation of the scenario for scatterplot A and some for scatterplot B...no clear pattern for the group as a whole.

NOTE: There are comments with "HINTS" in the cells requiring the formulas to calculate central tendency.

	1st attempt	2nd attempt
mean		<i>HINT: = average(highlight cells containing data)</i>
median		
mode		

Facilitation Questions

- What do mean, median, and mode describe about any set of data?

Answers may vary...

Mean – the value of each data point should all data points be "evened out"

Median – the value of the data point in the "middle" when considering the data points in numerical order (one-half are equal or greater than the median and one-half are equal or less than the median)

Mode – the value of the data point that occurs more often than other data points

- Looking at your data, how do the mean, median, and mode for the 1st attempt compare to that in the 2nd attempt? What might this imply about the comparison of the number of revolutions in the 2nd attempt as related to those in the 1st attempt? *Answers may vary.*
- Have the students share their answer to #9. What are you looking for in the data when you try to determine the measure of central tendency that will best describe the data?

Answers may vary.

Mean – data is clustered with no outliers

Median – most of the data is clustered except for one or more outliers

Mode – if one piece of data appeared significantly more times than others

EXPLAIN

The Explain portion of the lesson is directed by the teacher to allow the students to formalize their understanding of the TEKS addressed in the lesson.

- Once the students have completed their work on the spreadsheet **RoundandRound**, display **Transparency 3** to debrief. (Transparency 3 contains the same graphs that are on the RoundandRound spreadsheet.)
- Guide the students in drawing trendlines (if possible) on the three original scatterplots using the spreadsheet's drawing toolbar.

Facilitation Questions

For each scatterplot (A, B, C)

- Use the drawing toolbar in the spreadsheet to draw a line that would include the data points if the number of revolutions on the second attempt for each student were the same as the first attempt. ($y=x$) Does this line "fit" the data? Why or why not?

Answers may vary. A line $y = kx$ (proportional) will best "fit" to scatterplot A because it follows a similar trend...as x increases, y increases. Line $y = kx$ does not "fit" with scatterplots B or C because the data on those scatterplots does not fit a similar trend.

- For each scatterplot on the spreadsheet, is it possible to click on the red line below the scatterplot and place it on the scatterplot in such a way that it better exemplifies the relationships/trends in the data? (Click on the red line to move it and select "draw," "rotate," and "free rotate" to rotate the line.)
Answers may vary. Minor adjustments (as compared to line $y = x$) may be made for scatterplot A, whereas the red trend lines for scatterplots B and C should be significantly different than line $y = kx$.

- Select students to draw possible trendlines on the transparency. Discuss any differences in opinion. Use the trendlines to make predictions.

Facilitation Questions

- After drawing a trendline for scatterplot A, consider the points that would fall on or near the trendline we drew. As the number of revolutions made on the 1st attempt increases, what happens to the corresponding number of revolutions made on the 2nd attempt?
They increase as well.
- What type of correlation (trend) is this?
A positive correlation (trend)
- Based on this trendline, about how many revolutions would you expect students to make on the second attempt if they made 13 revolutions on their 1st attempt? What about if they had made 30 revolutions? 50 revolutions?
Answers may vary slightly depending on how the trendline was drawn.
- After drawing a trendline for scatterplot B, consider the points that would fall on or near the trendline we drew. As the number of revolutions made on the 1st attempt increases, what happens to the corresponding number of revolutions made on the 2nd attempt?
They increase.
- What type of correlation (trend) is this?
A positive correlation (trend)
- Based on this trendline, about how many revolutions would you expect students to make on their second attempt if they made 13 revolutions on their 1st attempt? What about if they had made 30 revolutions? 50 revolutions?
Answers may vary slightly depending on how the trendline was drawn.
- After attempting to draw a trendline for scatterplot C, why is it more difficult to draw a trendline on this scatterplot?
Answers may vary. Students should note that there is no clear pattern in the number of revolutions on the 2nd attempt (increasing or decreasing) as the number of revolutions increases on the 1st attempt. Visually, the points do not cluster around any line, rather they are spread more randomly throughout the scatterplot.
- What type of correlation (trend) is this?
There is no correlation (trend)
- Knowing we did not draw a trendline, about how many revolutions would you expect students to make on the second attempt if they made 13 revolutions on their 1st attempt? What about if they had made 30 revolutions? 50 revolutions?
With no clear trend, it is impossible to make a prediction based on this data alone.

4. Use **Transparency 4** to debrief the data in relationship to the mean or median.

5. Draw in the mean lines and discuss the characteristics (in relationship to the mean) of the pieces of data in each of the four resulting quadrants. (Students can use the drawing toolbar to draw the lines on their spreadsheet as you draw them on the transparency.)

Facilitation Questions

- Another way of looking at the data, other than a trendline, is to look at it in relationship to a central tendency such as mean or median. Look at scatterplot C where it was difficult to draw a trendline. What is the mean(average) number of revolutions made on the 1st attempt?
8.6667 (Draw in a vertical line at 8.6667 on the x-axis.)
- About how many data points fell below the mean? above the mean? What does this say about the data from those students?
6 below...3 above... Rationale may vary...should include a discussion about outliers and/or the spread of the data.
- What is the mean(average) number of revolutions made on the 2nd attempt?
10.44 (Draw in a horizontal line at 10.44 on the y-axis.)
- About how many data points fell below that mean? Above the mean? What does this say about the data from those students?
4 below...5 above... Rationale may vary...should include a discussion about outliers and/or the spread of the data.
- When you look at both mean lines, the data points divide into 4 groups. Describe the characteristic of each group.
Below the average on both attempts, below the average on the 1st attempt and above the average on the 2nd, above the average on the 1st attempt and below the average on the 2nd, above average on both attempts
- Why do you think the number of data points varies from group to group?
The lines were drawn using the mean values. Outliers may "pull" the line away from the center.

6. Use the second copy of Transparency 4 to draw in the median lines and discuss the characteristics (in relationship to the mean) of the pieces of data in each of the four resulting quadrants. (Students can use the drawing toolbar to draw the lines on their spreadsheet as you draw them on the transparency.)

Facilitation Questions

- If you were to do the same for the median lines, how do you think the data will be spread among the 4 groups?
Because the medians are the midpoints, the data should be evenly spread between the groups.
- Why might you want to look at the data in this manner?
Answers may vary. This is another way (other than trend lines) to communicate the relationship between the number of revolutions on the corresponding 1st and 2nd attempts.

ELABORATE

The Elaborate portion of the lesson provides an opportunity for the student to apply the concepts of the TEKS within a new situation. This part of the lesson is designed for groups of 2 students or individual investigation.

1. Direct students (or pairs of students if there is not enough technology available) to open the spreadsheet **BabyName**. Have students record their answers on notebook paper if the option of printing their work is not available.
2. Read the "Given" and the "Question" and have students turn to a neighbor and share their thoughts before sharing with the large group.

Facilitation Questions

- Read the "Given" statements and tell me what that means in your own words.
Answers may vary. Look for paraphrasing that connects "popularity" of names with frequency and the concept of ranking.
- Why might the popularity of certain names vary over time?
Answers may vary. Students might consider factors such as culture, famous figures, etc.
- Read the "Question" and turn to your neighbor and share your thoughts.
(Pause) Do you think there will be a difference in the change in popularity of boy names versus girl names? Why or why not?
Answers may vary. Accept all answers for now.

3. Read through the directions. Make sure students can access the data website (<http://www.ssa.gov/OACT/babynames/>) or print and have hard copies of the data available if using the Internet is not an option. NOTE: Using technology to search on the Internet is much more efficient than searching on paper.
4. Have students complete spreadsheet.

Facilitation Questions

- (After locating the top ten boy names for 1965...) Look at the data source and explain again how certain names make this list and others do not.
Answers may vary. Students should determine that the data comes from counting the number of times a particular first name was put on applications for Social Security cards for newborns. Ex. Since "Michael" is ranked first, this means that there were more newborns with the first name of "Michael," according to the information parents gave on their baby's Social Security card application, than any other first name.

5. Use **Transparency 5** to debrief the activity.

Facilitation Questions

- What was the only central tendency not calculated in the activity?
Mode
- Why do you think mode was not included?
Since the names each have a unique rank (understanding that it would be highly unlikely that two names would occur exactly the same number of times), there will be no mode.
- When answering #9, what characteristics in the data made you choose to draw the lines for the mean or for the median?
Answers may vary. Looking for some discussion of clustering of data and/or outliers.
- (Transparency 5 – Draw in the lines based on the discussion from the previous question, then draw in a line representing the ranking from 1965.) Describe the relationship among the three lines.
Answers may vary. Students should note that the central tendency line for the rankings of boy names is much closer to the line representing the rankings in 1965...a visual demonstrating how the popularity of those boy names has remained somewhat steady in comparison to the girl names of similar rankings.
- Look at the lines you drew on the scatterplot. Would you say, based on the data you have, that you could better predict the change in popularity of a boy name or a girl name?
Answers may vary. Since the points representing boy names are more clustered around that line, chances are the popularity of a boy name will change less than that of girl names with similar original rankings.

Facilitation Questions

- Think back to your response to the “Question” at the beginning of the activity. Given the additional information you now have, do you need to revise or elaborate on your initial thoughts?

Answers may vary. In general, the selection of names for girls tends to be more susceptible to varying trends than names for boys. In other words, based on the data for the names we researched, the popularity of a particular name for a girl is much more likely to change over time; whereas the popularity for a particular name for a boy is more likely to remain somewhat steady. Just a thought... This could be related to the custom many have of designating males as the ones who will carry on the family name.

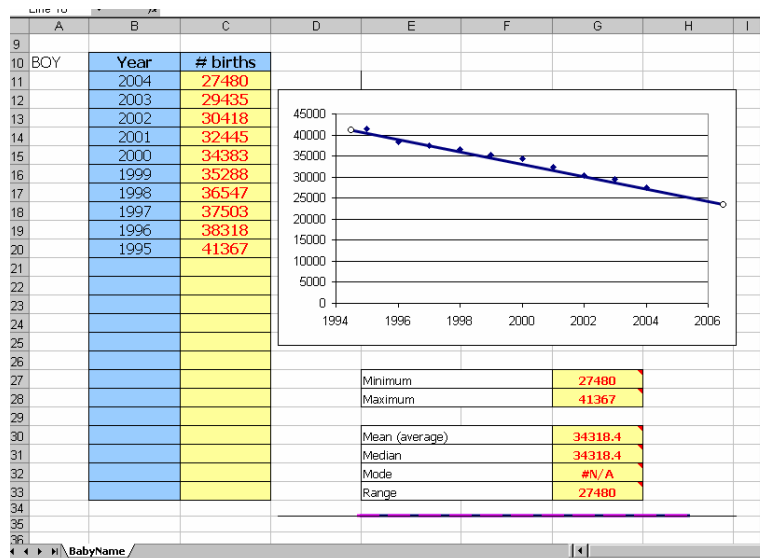
- How are the trends in this data similar to or different from the trend you saw in your data from the hula hoop experiment?

Answers may vary. For example, if there were no clear trend between the numbers of revolutions made on the 1st and 2nd attempts, students might see similarities between that data and the data for the girl names.

EVALUATE

The Evaluate portion of the lesson provides the student with an opportunity to demonstrate his or her understanding of the TEKS addressed in the lesson.

1. Distribute **What’s In A Name?** activity sheet to each student.
2. Upon completion of the **What’s In A Name?** activity sheet, the teacher should use a rubric to assess student understanding of the concepts addressed in this lesson.



Scatterplot Lesson Spreadsheet

Answer may vary somewhat...around 20,000 people with the top ranking boy name in 2010

Justifications may vary...Based on the data in this scatterplot, there is a negative correlation between the year and the number of people with the top ranking boy name. As the years increase, the number of people with the selected name decreases. If this trend were to continue, you would expect around 20,000 people with the top ranking boy name in 2010. Students could also justify their prediction with one of the statistical measures such as finding a range and extending it to future years.

Answers and Error Analysis for selected response questions:

Question Number	TEKS	Correct Answer	Conceptual Error	Conceptual Error	Procedural Error	Procedural Error	Guess
1	8.12(B)	A	B	C			D
2	8.12(B)	C	A	B			D
3	8.12(A)	C	A		B		D
4	8.12(B)	C	A	B			D

EXPLORE

Round and Round

Open the **RoundandRound** spreadsheet.

A. Input your class data from Transparency 2. (Use the table that starts in row 6.)

Sketch the resulting scatterplot.

Will vary depending on data

B. For each statement, choose the scatterplot(s) that best represents the situation.

 A, B, C 1. After the 1st attempt, most students were able to increase the number of revolutions on their 2nd attempt.

 none 2. After the 1st attempt, most students made fewer revolutions on their 2nd attempt.

 A 3. The number of revolutions on the 1st attempt is about the same as the number of revolutions on the 2nd attempt.

 C 4. There is not a strong relationship between the number of revolutions made in the two attempts.

 B 5. Most students did considerably better on their 2nd attempt than on their 1st attempt.

 vary 6. Based on the data you have from your class, which scatterplot would look most like yours? Explain.

7. Now that you have analyzed possible scenarios for scatterplots A, B, and C, write a statement that describes the relationship between the 1st attempt and 2nd attempt for your class.

Answers may vary.

8. Use formulas to calculate the mean, median, and mode of the data for the 1st attempt and for the 2nd attempt. (Use the table in rows 50-53.) Record the results below.

Answers may vary.

	1st attempt	2nd attempt
mean		
median		
mode		

9. Which measure of central tendency best describes the number of revolutions made on the 1st attempt and 2nd attempt? Explain your choice.

Answers may vary.

Transparency 1

How many revolutions of a hula hoop can you achieve in one attempt?

Will this prediction change for a second attempt? Why or why not?

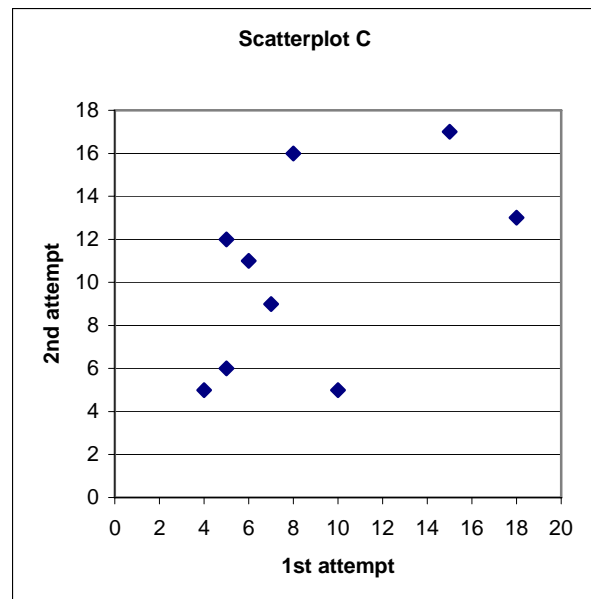
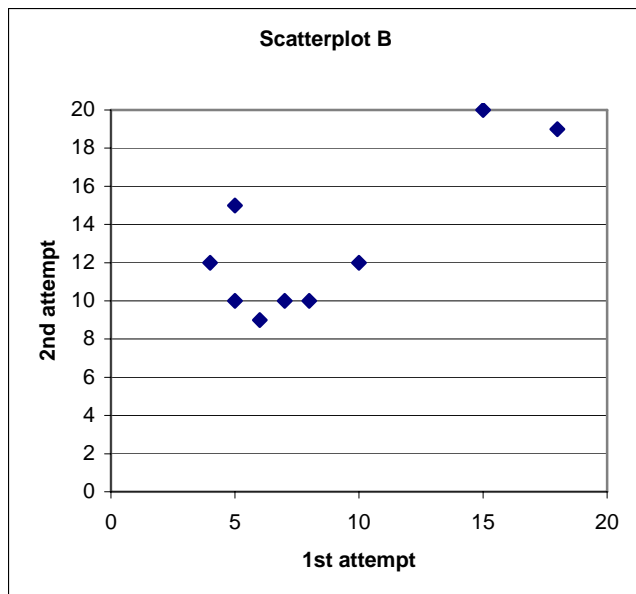
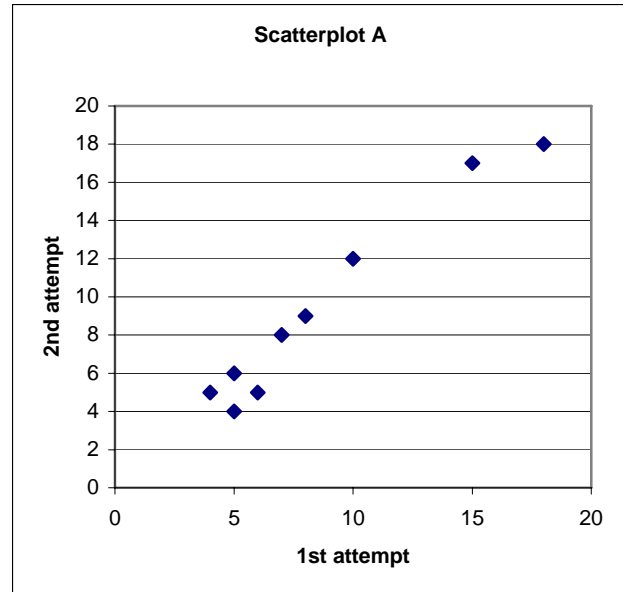
How might we gather data to test our predictions?

Transparency 2

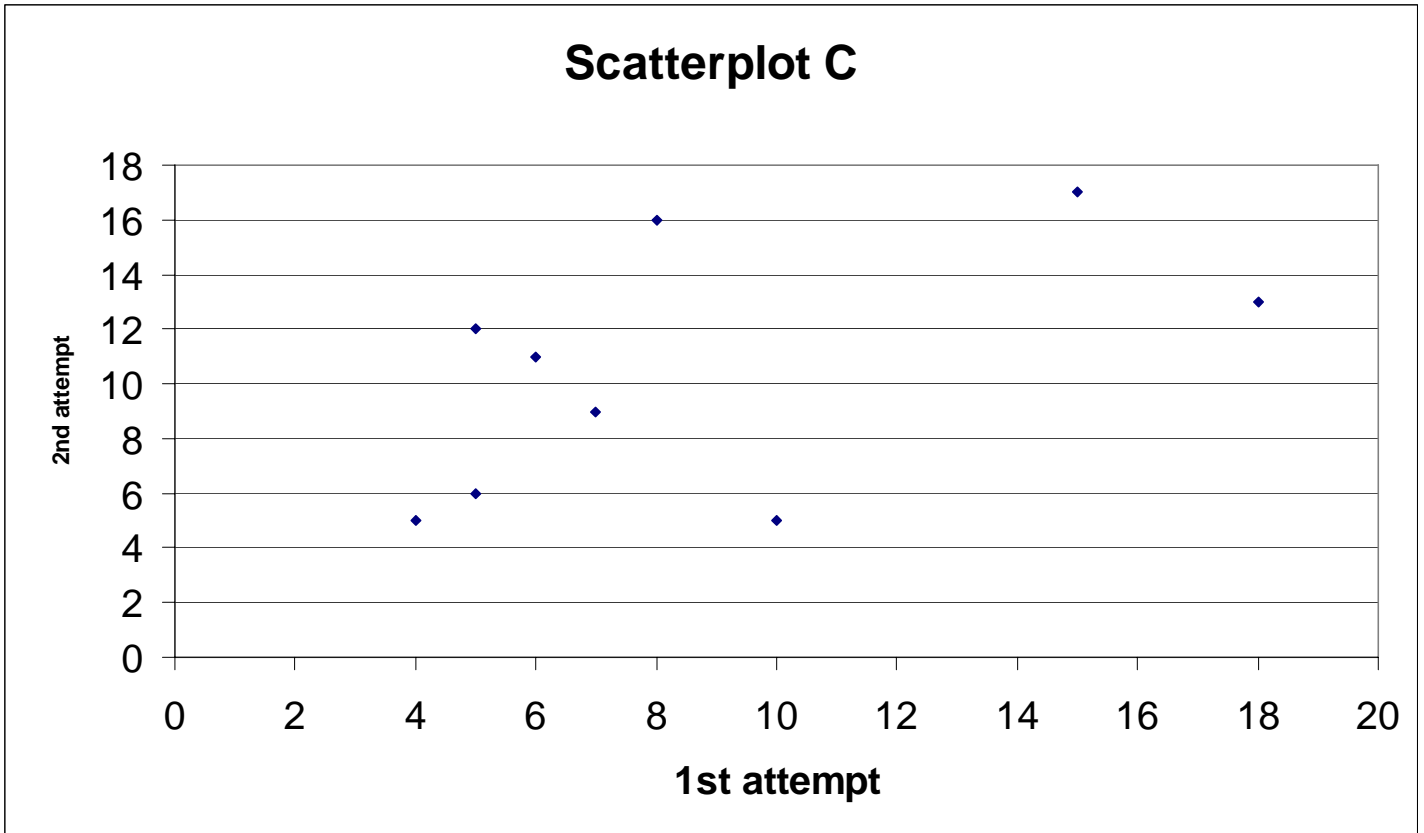
Student	# Revolutions on 1 st Attempt	# Revolutions on 2 nd Attempt

Transparency 3

1. If possible, sketch a trendline.
2. Predict the number of revolutions on the 2nd attempt if the number on the 1st attempt was 13...30...100.

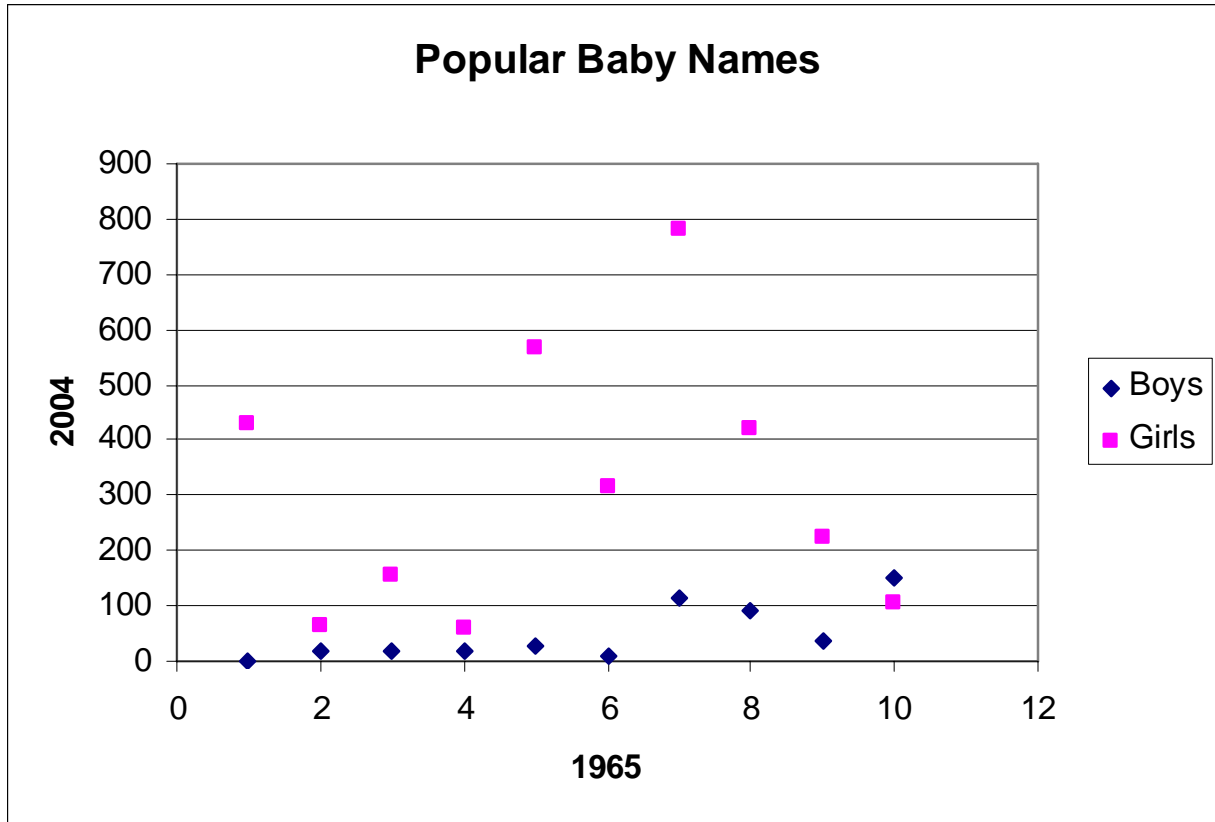


Transparency 4



	1 st attempt	2 nd attempt
mean	8.6667	10.44
median	9.0741	10.94

Transparency 5



	1965	2004 – Boys	2004 – Girls
	— — —	- - - - -
Mean	5.5	48.1	312.1
Median	5.5	23.5	269.5
Range	9	147	504

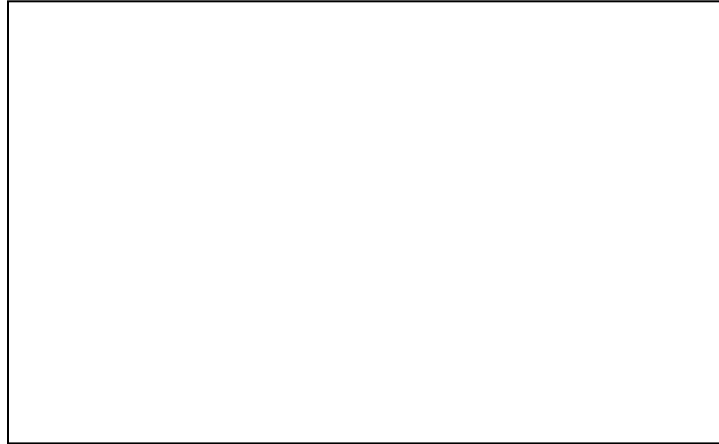
Student Name(s) _____ Date _____

Round and Round

Open the **RoundandRound** spreadsheet.

A. Input your class data from Transparency 2. (Use the table that starts in row 6.)

Sketch the scatterplot that resulted from entering class data.



B. For each statement, choose the scatterplot(s) that best represents the situation.

_____ 1. After the 1st attempt, most students were able to increase the number of revolutions on their 2nd attempt.

_____ 2. After the 1st attempt, most students made fewer revolutions on their 2nd attempt.

_____ 3. The number of revolutions on the 1st attempt is about the same as the number of revolutions on the 2nd attempt.

_____ 4. There is not a strong relationship between the number of revolutions made in the two attempts.

_____ 5. Most students did considerably better on their 2nd attempt than on their 1st attempt.

_____ 6. Based on the data you have from your class, which scatterplot would look most like yours? Explain.

7. Now that you have analyzed possible scenarios for scatterplots A, B, and C, write a statement that describes the relationship between the 1st attempt and 2nd attempt for your class.

8. Use formulas to calculate the mean, median, and mode of the data for the 1st attempt and for the 2nd attempt. (Use the table in rows 50-53.) Record the results below.

	1st attempt	2nd attempt
mean		
median		
mode		

9. Which measure of central tendency best describes the number of revolutions made on the 1st attempt and 2nd attempt? Explain your choice.

Student Name(s) _____ Date _____

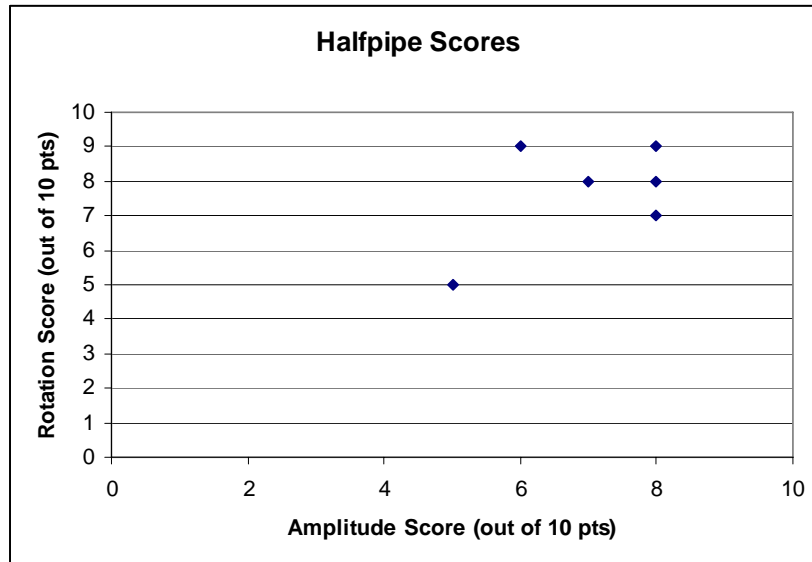
What's In A Name?

Access the website <http://www.ssa.gov/OACT/babynames/>. In an earlier activity we compared the ranking of the top ten names of your parents' generation (1965) to the ranking of those names today to answer the question about how the popularity of names stands the test of time.

- a. Consider the following set of questions.
How has the number of people having the most popular boy name changed over the last 10 years? How many people do you predict might have the most popular name in 2010?
- b. Open the **WhatName** spreadsheet and input the data to create a scatterplot.
- c. Draw a trendline if appropriate.
- d. Calculate the mean, median, and range of your data.
- e. Respond to the questions in part a. Justify your answers using the scatterplot, trendline, and/or statistical measurements to support your conclusions.

Student Name(s) _____ Date _____

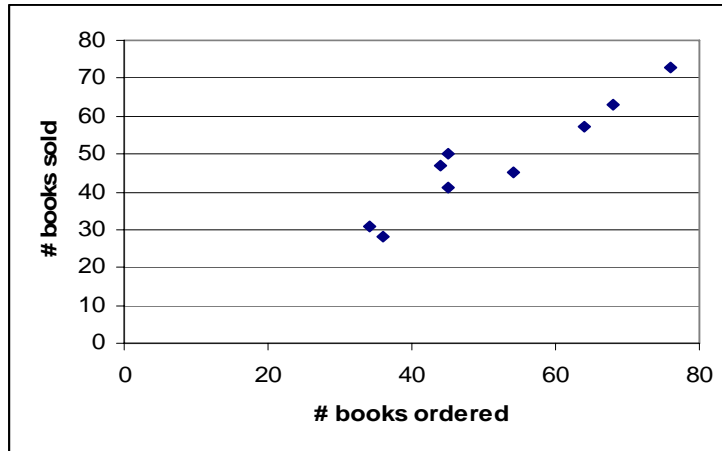
1. The scatterplot below compares the score for amplitude (height) to the score for rotations (spins and flips) for six skateboarders at the weekend meet.



- Which of the following statements would be supported by the scatterplot?
- A. As the score for amplitude increases, the score for rotations tends to increase.
 - B. As the score for amplitude increases, the score for rotations tends to decrease.
 - C. As the score for rotations increases, the score for amplitude tends to decrease.
 - D. The score for rotations tends to be the same as the score for amplitude.

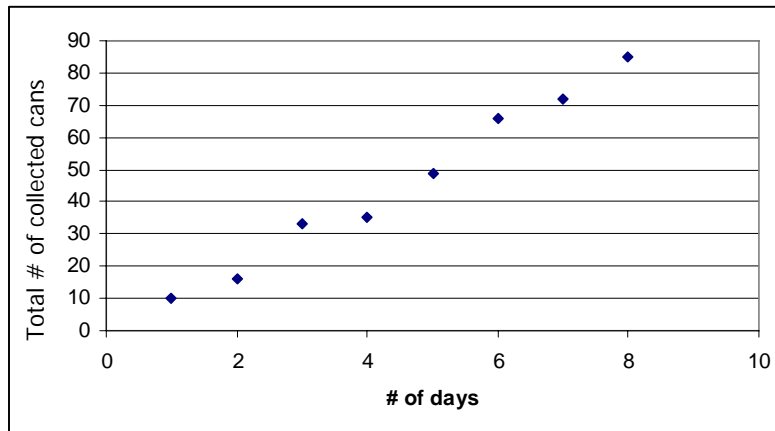
2. Which relationship, when graphed on a scatterplot, would NOT be described as having a positive trend?
- A. The number of fans in a football stadium compared to the noise level of the stadium.
 - B. The amount of money earned babysitting compared to the number of hours spent babysitting.
 - C. The number of miles driven compared to the amount of gasoline in the tank.
 - D. All of the above relationships have a positive trend.

3. The following scatterplot compares the number of books ordered through the school fund raiser to the number of books that were sold.



If the mean number of books ordered is about 52, estimate the mean number of books sold based on the trends in data in the scatterplot.

- A. greater than 55
 - B. between 50 and 55
 - C. between 45 and 50
 - D. less than 40
4. Ms. Smith's class is recording data about an aluminum can recycling project as shown in the scatterplot below.



At this rate, about how many days will it take to collect 150 cans?

- A. 150 days
- B. 20 days
- C. 15 days
- D. 10 days

Mathematics

8.12 The student uses statistical procedures to describe data. The student is expected to:

- (A) select the appropriate measure of central tendency or range to describe a set of data and justify the choice for a particular situation.
- (B) draw conclusions and make predictions by analyzing trends in scatterplots.
- (C) Select and use an appropriate representation for presenting and displaying relationships among collected data, including line plots, line graphs, stem and leaf plots, circle graphs, bar graphs, box and whisker plots, histograms, and Venn diagrams, with and without the use of technology.

Technology Applications

The student demonstrates knowledge and appropriate use of hardware components, software programs, and their connections. The student is expected to:

- (1)(a) demonstrate knowledge and appropriate use of operating systems, software applications, and communicate and networking components.
- (1)(c) demonstrate the ability to select and use software for a defined task according to quality, appropriateness, effectiveness, and efficiency.
- (1)(f) perform basic software application function including, but not limited to, opening an application program and creating, modifying, printing, and saving documents.
- (1)(h) use terminology related to the Internet appropriately including, but not limited to, electronic mail (e-mail), Uniform Resource Locators (URLs), electronic bookmarks, local area networks (LANs), wide area networks (WANs), World Wide Web (WWW) page, and HyperText Markup Language (HTML).

The student uses a variety of strategies to acquire information from electronic resources, with appropriate supervision. The student is expected to:

- (4)(a) Use strategies to locate and acquire desired information on LANs and WANs, including the Internet, intranet, and collaborative software.

Materials

Advanced Preparation:

- Internet access to <http://www.ssa.gov/OACT/babynames/>
- Copies of **Round and Round**, **Baby Names**, and **What's In A Name** worksheets for each student
- Access to a **TI-73** for each student or pair of students

For whole class demonstration:

- Transparencies 1 - 5
- Several hula hoops (up to 1 per 2 students to save time)

- Chart paper, markers

For each student:

- **Round and Round** worksheet
- **BabyNames** worksheet
- **What's In A Name** worksheet
- TI-73

ENGAGE

The Engage portion of the lesson is designed to create student interest in the concepts addressed. This part of the lesson is designed for whole class discussion/ demonstration.

1. Prompt a student to read the “Problem” aloud from **Transparency 1** and ask students to individually consider their prediction.
2. Prompt students to share their predictions with a neighbor before getting responses from the large group.

Facilitation Questions

- What data are you collecting and comparing for each student?
We are looking for, and comparing, the number of revolutions on the first attempt and the number of revolutions on the second attempt.
- Once several students have recorded their predictions on the numbers of revolutions on the first attempt, what will you need to consider when making predictions about the second attempt?
*Answers may vary, but you are looking for the fact that students recognize that, chances are, the change in the number of revolutions between attempts will not be exactly the same for all students—some will increase, some will not. Instead, they should look for trends that describe the change, such as “The number of revolutions for the second attempt **tends** to be _____.”*
- Will your prediction be true for every student? Why or why not?
No. Point out that this exemplifies how “real” data often does not fit into “clean” or exact patterns such as linear patterns. Instead, we have to look for any trends in the data.
- How many pieces of data would you need to make a prediction about the number of revolutions on a second attempt?
Answers may vary, but you are looking for students to recognize that the more data you have, the better defined any correlation will be.

3. Show the table on **Transparency 2** that will be used to collect the data and agree on a sample size of 10 to 15 students. You may want to agree ahead of time as to whether or not each student will be allowed a practice attempt. (If hula-hoops are not available, you can substitute another event such as paddle-ball, trash can

basketball, etc. The goal is to choose an event where the trend in the data is not obvious or that you could make a case for multiple trends. For example, one might make the case that students would have more revolutions with the hula-hoop on their 2nd attempt because of the practice they got during the 1st attempt or that they would have fewer on their 2nd attempt because they were tired from the 1st attempt or that there would be no clear correlation.)

4. Pair the students. One will hula-hoop twice and the other will record the number of revolutions on the 1st and 2nd attempt on **Transparency 2**.

Facilitation Questions

- If a student is able to complete 6 revolutions on the first attempt, what could happen on the second attempt? Why?
The number of revolutions could be more than 6, less than 6, or the same as 6 on the second attempt.
- If that same student were able to make 10 revolutions on the second attempt (an increase of 4 over the first attempt), would this mean that the same will be true for the next student? Why?
Possibly, but possibly not—we do not have enough data to make that prediction yet.

EXPLORE

The Explore portion of the lesson provides the student with an opportunity to be actively involved in the exploration of the mathematical concepts addressed. This part of the lesson is designed for groups of 2 students or individual investigation.

1. Distribute a TI-73 to each student (or pairs of students if there is not enough technology available).
2. Distribute and go over the directions on the **Round and Round** worksheet.
3. Direct the students to input their student data from **Transparency 2** in List 1 and List 2.

Facilitation Questions

- What will you put in list 1 [L1] and what will you put in list 2 [L2]?
List 1 will contain values from the data on the 1st attempt, and list 2 will contain values from the data on the 2nd attempt.
- Is the order in which you input the values important? Why or why not?
The order in which the pairs of data are input is not important, but keeping the pairs of values within the same row on the list is important. In other words, it is acceptable to input the data from student #2 before the data from student #1, but it is not acceptable to pair the value of the 1st attempt from student #1 with the value of the 2nd attempt from student #2.

4. Create a scatterplot. Press $\boxed{2\text{nd}}\boxed{Y=}$ to access the Plot menu. Press $\boxed{\text{ENTER}}$ to set the plot as shown below.



5. Press $\boxed{\text{WINDOW}}$ to set an appropriate window that would contain the data.

Facilitation Questions

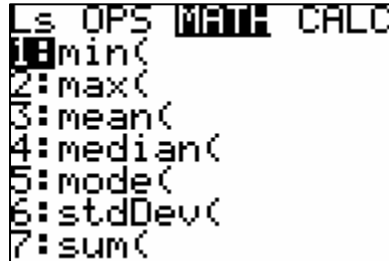
- Based on our data, what will you need to consider when setting the Xmin and Xmax in the window?
The Xmin will need to be as small or smaller than the least value in list 1 and the Xmax will need to be as large or larger than the greatest value in list 1.
- Based on our data, what will you need to consider when setting the Ymin and Ymax in the window?
The Ymin will need to be as small or smaller than the least value in list 2 and the Ymax will need to be as large or larger than the greatest value in list 2.
- How would you describe any trends that you might see in the scatterplot drawn from the data from your class?
Answers may vary depending on the data collected. Look for statements such as "As the number of revolutions during the 1st attempt increases, the number of revolutions during the 2nd attempt (increases/decreases)." Or "Students tend to _____."

6. Point out the three scatterplots on the **Transparency 3**. Make sure the students understand that these were drawn based on fictitious data, and not their own, as they answer questions #1 – 7 on the worksheet.

Facilitation Questions

- Describe a possible scenario that would produce each of the three scatterplots.
Answers may vary...
Scatterplot A – Due to their practice in attempt 1, students were able to make more revolutions in attempt 2.
Scatterplot B – Due to their efforts in attempt 1, students did not have as much energy to make as many revolutions in attempt 2.
Scatterplot C – Some students followed the explanation of the scenario for scatterplot A and some for scatterplot B...no clear pattern for the group as a whole.

7. Prompt the students to calculate and compare the measures of central tendencies, including mean, median, and mode (worksheet #8 and 9).
 - a. Go to the home screen.
 - b. Press $\boxed{2\text{nd}}\boxed{\text{LIST}}$ then arrow over to the Math menu. Choose the appropriate measure.



- c. Press $\boxed{\text{ENTER}}$. Then press $\boxed{2\text{nd}}\boxed{\text{LIST}}$ to choose the appropriate list. Press $\boxed{\text{ENTER}}$.
- d. Record on chart paper.

	List 1	List 2
Mean		
Median		
Mode		

Facilitation Questions

- What do mean, median, and mode describe about any set of data?
Answers may vary...
Mean – the value of each data point should all data points be “evened out”
Median – the value of the data point in the “middle” when considering the data points in numerical order (one-half are equal or greater than the median and one-half are equal or less than the median)
Mode – the value of the data point that occurs more often than other data points
- Looking at your data, how do the mean, median, and mode for the 1st attempt compare to that in the 2nd attempt? What might this imply about the comparison of the number of revolutions in the 2nd attempt as related to those in the 1st attempt?
Answers may vary.
- Have the students share their answer to #9. What are you looking for in the data when you try to determine the measure of central tendency that will best describe the data?
Answers may vary.
Mean – data is clustered with no outliers
Median – most of the data is clustered except for one or more outliers
Mode – if one piece of data appeared significantly more times than others

EXPLAIN

The Explain portion of the lesson is directed by the teacher to allow the students to formalize their understanding of the TEKS addressed in the lesson.

1. Once the students have completed their work, put up **Transparency 3** to debrief.
2. Guide the students in drawing trendlines (if possible) on the three original scatterplots. Select students to draw possible trendlines on the transparency. Discuss any differences in opinion. Use the trendlines to make predictions.

Facilitation Questions

For each scatterplot (A, B, C)

- Draw a line that would include the data points if the number of revolutions on the second attempt for each student was the same as their first attempt. ($y=x$) Does this line "fit" the data? Why or why not?

Answers may vary. Line $y = x$ will best "fit" to scatterplot A because it follows a similar trend...as x increases, y increases. Line $y = x$ does not "fit" with scatterplots B or C because the data on those scatterplots does not fit a similar trend...as x increases, y increases.

- For each scatterplot, is it possible to draw a line on the scatterplot in such a way that it better exemplifies the relationships/trends in the data?

Answers may vary. Minor adjustments (as compared to line $y = x$) may be made for scatterplot A, whereas the trend lines for scatterplots B and C should be significantly different than line $y = x$.

- (After drawing a trendline for scatterplot A) Consider the points that would fall on or near the trendline we drew. As the number of revolutions made on the 1st attempt increases, what happens to the corresponding number of revolutions made on the 2nd attempt? *They increase as well.*

- What type of correlation (trend) is this?

A positive correlation (trend)

- Based on this trendline, about how many revolutions would you expect students to make on their second attempt if they made 13 revolutions on their 1st attempt? What about if they had made 30 revolutions? 50 revolutions?

Answers may vary slightly depending on how the trendline was drawn.

- (After drawing a trendline for scatterplot B) Consider the points that would fall on or near the trendline we drew. As the number of revolutions made on the 1st attempt increases, what happens to the corresponding number of revolutions made on the 2nd attempt? *They increase.*

- What type of correlation(trend) is this?

A positive correlation(trend).

- Based on this trendline, about how many revolutions would you expect students to make on their second attempt if they made 13 revolutions on their 1st attempt? What about if they had made 30 revolutions? 50 revolutions?

Answers may vary slightly depending on how the trendline was drawn.

Facilitation Questions

- (After attempting to draw a trendline for scatterplot C) Why is it more difficult to draw a trendline on this scatterplot?
Answers may vary. Students should note that there is no clear pattern in the number of revolutions on the 2nd attempt (increasing or decreasing) as the number of revolutions increases on the 1st attempt. Visually, the points do not cluster around any line, rather they are spread more randomly throughout the scatterplot.
- What type of correlation (trend) is this?
There is no correlation (trend).
- Knowing we did not draw a trendline, about how many revolutions would you expect students to make on their second attempt if they made 13 revolutions on their 1st attempt? What about if they had made 30 revolutions? 50 revolutions?
With no clear trend, it is impossible to make a prediction based on this data alone.

3. Use **Transparency 4** to discuss looking at the data in relationship to the means or medians.
4. Draw in the mean lines and discuss the characteristics (in relationship to the mean) of the pieces of data in each of the four resulting quadrants.

Facilitation Questions

- Another way of looking at the data, other than a trendline, is to look at it in relationship to a central tendency such as mean or median. Look at scatterplot C where it was difficult to draw a trendline. What is the mean(average) number of revolutions made on the 1st attempt?
8.6667 (Draw in a vertical line at 8.6667 on the x-axis.)
- About how many data points fell below the mean? above the mean? What does this say about the data from those students?
6 below...3 above... Rationale may vary...should include a discussion about outliers and/or the spread of the data.
- What is the mean(average) number of revolutions made on the 2nd attempt?
10.44 (Draw in a horizontal line at 10.44 on the y-axis.)
- About how many data points fell below that mean? Above the mean? What does this say about the data from those students?
4 below...5 above... Rationale may vary...should include a discussion about outliers and/or the spread of the data.

Facilitation Questions

- When you look at both mean lines, the data points divide into 4 groups. Describe the characteristics of each group.
Below the average on both attempts, below the average on the 1st attempt and above the average on the 2nd, above the average on the 1st attempt and below the average on the 2nd, above average on both attempts
- Why do you think the number of data points varies from group to group?
The lines were drawn using the mean values. Outliers may "pull" the line away from the center.

5. Draw in the median lines and discuss the characteristics (in relationship to the mean) of the pieces of data in each of the four resulting quadrants.

Facilitation Questions

- If you were to do the same for the median lines, how do you think the data will be spread between the 4 groups?
Because the medians are the midpoints, the data should be evenly spread between the groups.
- Why might you want to look at the data in this manner?
Answers may vary. This is another way (other than trend lines) to communicate the relationship between the number of revolutions on the corresponding 1st and 2nd attempts.

ELABORATE

*The Elaborate portion of the lesson provides an opportunity for the student to apply the concepts of the TEKS within a new situation. This part of the lesson is designed for **groups of 2 students or individual investigation.***

1. Distribute a **Baby Name** worksheet to each student.
2. Read the "Given" and the "Question" and have students turn to a neighbor and share their thoughts before sharing with the large group.

Facilitation Questions

- Read the "Given" statements and tell me what that means in your own words.
Answers may vary. Look for paraphrasing that connects "popularity" of names with frequency and the concept of ranking.
- Why might the popularity of certain names vary over time?
Answers may vary. Students might consider factors such as culture, famous figures, etc.
- Read the "Question" and turn to your neighbor and share your thoughts.
(Pause) Do you think there will be a difference in the change in popularity of boy names versus girl names? Why or why not?
Answers may vary. Accept all answers for now.

3. Read through the directions and make sure students can access the data website (<http://www.ssa.gov/OACT/babynames/>) or print and have hard copies of the data available if using the Internet is not an option. NOTE: Using technology to search on the Internet is much more efficient than searching on paper.
4. Begin completing the table for the boy names as a large group to ensure students are comfortable with accessing the appropriate data.

Facilitation Questions

- (After locating the top ten boy names for 1965...) Look at the data source and explain again how certain names make this list and others do not.
Answers may vary. Students should determine that the data comes from counting the number of times a particular first name was put on applications for Social Security cards for newborns. Ex. Since "Michael" is ranked first, this means that there were more newborns with the first name of "Michael," according to the information parents gave on their baby's Social Security card application, than any other first name.

5. Direct the students to create a scatterplot as before. Once the scatterplot is complete, direct the students' attention to the questions for the scatterplot which will have them calculate and interpret the mean, median, and range of the data. Use **Transparency 5** to debrief the activity.

Facilitation Questions

- (Upon completion of the scatterplot...) Have students share their answers to the questions below the scatterplot.
Answers may vary..
- What was the only central tendency not calculated in the activity?
Mode
- Why do you think mode was not included?
Since the names each have a unique rank (understanding that it would be highly unlikely that two names would occur exactly the same number of times), there will be no mode.
- When answering #9, what characteristics in the data made you choose to draw the lines for the mean or for the median?
Answers may vary. Look for some discussion of clustering of data and/or outliers.
- (Transparency 5 – Draw in the lines based on the discussion from the previous question, then draw in a line representing the ranking from 1965.) Describe the relationship among the three lines.
Answers may vary. Students should note that the central tendency line for the rankings of boy names is much closer to the line representing the rankings in 1965...a visual demonstrating how the popularity of those boy names has remained somewhat steady in comparison to the girl names of similar rankings.
- Look at the lines you drew on the scatterplot. Would you say, based on the data you have, that you could better predict the change in popularity of a boy name or a girl name?
Answers may vary. Since the points representing boy names are more clustered around that line, chances are the popularity of a boy name will change less than that of girl names with similar original rankings.
- Think back to your response to the "Question" at the beginning of the activity. Given the additional information you now have, do you need to revise or elaborate on your initial thoughts?
Answers may vary. In general, the selection of names for girls tends to be more susceptible to varying trends than names for boys. In other words, based on the data for the names we researched, the popularity of a particular name for a girl is much more likely to change over time; whereas the popularity for a particular name for a boy is more likely to remain somewhat steady. Just a thought...This could be related to the custom many have of designating males as the ones who will carry on the family name.
- How are the trends in this data similar to or different than the any trend you saw in your data from the hula hoop experiment?
Answers may vary. For example, if there was no clear trend between the numbers of revolutions made on the 1st and 2nd attempts, students might see similarities between that data and the data for the girl names.

EVALUATE

The Evaluate portion of the lesson provides the student with an opportunity to demonstrate his or her understanding of the TEKS addressed in the lesson.

1. Distribute **What's In A Name?** activity sheet to each student.
2. Upon completion of the **What's In A Name?** activity sheet, the teacher should use a rubric to assess student understanding of the concepts addressed in this lesson.

Answers may vary somewhat...around 20,000 people with the top ranking boy name in 2010

Justifications may vary...Based on the data in this scatterplot, there is a negative correlation between the year and the number of people with the top ranking boy name. As the years increase, the number of people with the selected name decreases. If this trend was to continue, you would expect around 20,000 people with the top ranking boy name in 2010. Students could also justify their prediction with one of the statistical measures such as finding a range and extending it to future years.

Answers and Error Analysis for selected response questions:

Question Number	TEKS	Correct Answer	Conceptual Error	Conceptual Error	Procedural Error	Procedural Error	Guess
1	8.12(B)	A	B	C			D
2	8.12(B)	C	A	B			D
3	8.12(A)	B	A	D	C		D
4	8.12(B)	A	D		B		C

Round and Round

A. Input your class data from Transparency 2.

Sketch the resulting scatterplot.

Will vary depending on data

B. For each statement, choose the scatterplot(s) that best represents the situation.

 A, B, C 1. After the 1st attempt, most students were able to increase the number of revolutions on their 2nd attempt.

 none 2. After the 1st attempt, most students made fewer revolutions on their 2nd attempt.

 A 3. The number of revolutions on the 1st attempt is about the same as the number of revolutions on the 2nd attempt.

 C 4. There is not a strong relationship between the number of revolutions made in the two attempts.

 B 5. Most students did considerably better on their 2nd attempt than on their 1st attempt.

 vary 6. Based on the data you have from your class, which scatterplot would look most like yours? Explain.

Round and Round (continued)

7. Now that you have analyzed possible scenarios for scatterplots A, B, and C, write a statement that describes the relationship between the 1st attempt and 2nd attempt for your class.

Answers may vary.

8. Use formulas to calculate the mean, median, and mode of the data for the 1st attempt and for the 2nd attempt. Record the results below.

Answers may vary.

	1st attempt	2nd attempt
mean		
median		
mode		

9. Which measure of central tendency best describes the number of revolutions made on the 1st attempt and 2nd attempt? Explain your choice.

Answers may vary.

Baby Names

(based on data from Social Security card applications)

Given: Some baby names are more popular (occur more often) than others. The list of the most popular baby names changes from year to year. While some names are used less over time, others remain popular.

Question: Over the last 40 years, do you think boy names or girl names have been less "trendy"? In other words, do you think children in your generation are more likely to have the same names as adults in your parents' generation if they are boys or girls? Today you will research to compare the ranking of the most popular names in the year 1965 to the ranking of those names in the year 2004 (a span of 40 years).

Directions:

1. Go to the website below to determine the top 10 names for boys and girls in the year 1965.
2. Record the names missing in the tables below.
3. Look up ranks missing for each name for the year 2004 and add that data to the chart. (The database lists the top 1000 names. Use a rank of 1001 if a name is not included.)

<http://www.ssa.gov/OACT/babynames/>

BOYS		
Name	1965 Rank	2004 Rank
Michael	1	2
John	2	18
David	3	16
James	4	17
Robert	5	29
William	6	8
Mark	7	113
Richard	8	92
Thomas	9	37
Jeffrey	10	149

GIRLS		
Name	1965 Rank	2004 Rank
Lisa	1	431
Mary	2	63
Karen	3	154
Kimberly	4	61
Susan	5	565
Patricia	6	317
Donna	7	781
Linda	8	422
Cynthia	9	222
Angela	10	105

4. Looking at the data in the table, what do you notice about the change in rank of the top 10 names for boys versus the change in rank of the top 10 names for girls over the last 40 years?

Answers may vary. Students should note that the change in the rankings of the girls names is much larger than that of the boys.

5. Looking at the data in the scatterplot, explain how any observations you made from the table in #4 are reflected in the scatterplot.

Answers may vary. Students should note that the points representing the rankings of the girls names are much more spread apart (compared to their rank in 1965) than the boys.

6. Calculate the mean rank of the given boy names for 2004 and girl names for 2004. How many boy names and how many girl names were more popular than the mean rank?

7 boy names and 5 girl names are less than their respective means, therefore more popular.

7. Calculate the median rank of the given boy names for 2004 and girl names for 2004. How many boy names and how many girl names were more popular than the median rank?

5 boy names and 5 girl names are less than their respective medians, therefore more popular.

8. Calculate the range in the rankings of the given boy names for 2004 (cell B43) and girl names for 2004. How does the range value for the boys compare to that of the girls? What does this mean?

Answers may vary. Students should note that the range in 2004 rankings for the boys is much smaller than that of the girls, meaning their rankings are closer together in value.

9. Determine whether the mean or median best describes the data. Position the lines below the scatterplot (solid for boys and dashed for girls) on the scatterplot to represent these measurements. What does the position of the lines on the scatterplot emphasize about the relationship between how the popularity of the top ten names for boys and girls has changed from your parents' generation to now?

Answers may vary...a case could be made for the median being the better descriptor due to the wide spread in the data. In either case, students should note that any change in popularity over time was more dramatic for girl names than it was for boy names since that line is farther from the original rankings.

10. Describe the relationship between the points on the scatterplot and the lines you drew in #9 for the boy names and the girl names. What does this mean?

Answers may vary. The points representing rankings of the boy names are much closer to that line than the points representing rankings of the girl names are to their line.

Transparency 1

How many revolutions of a hula hoop can you achieve in one attempt?

Will this prediction change for a second attempt? Why or why not?

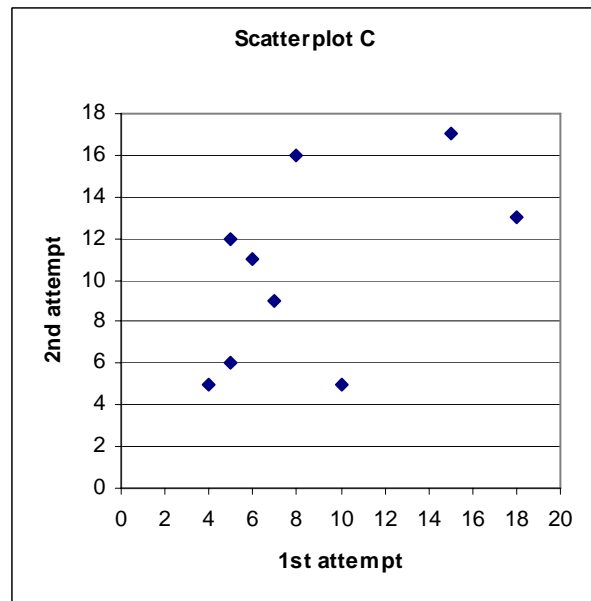
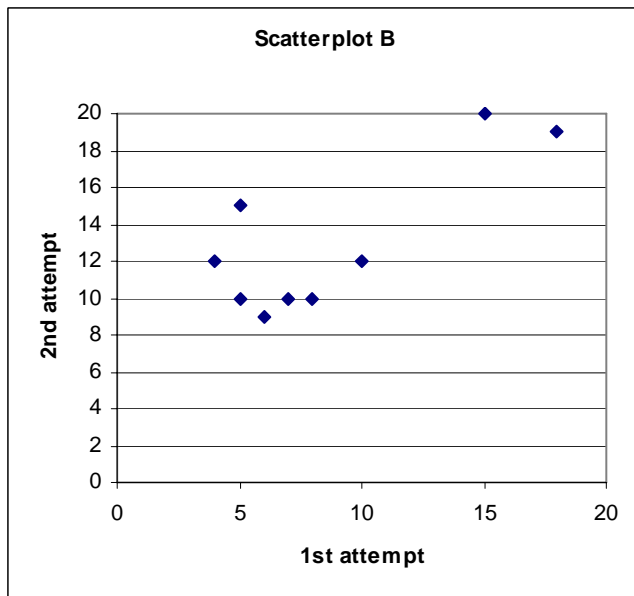
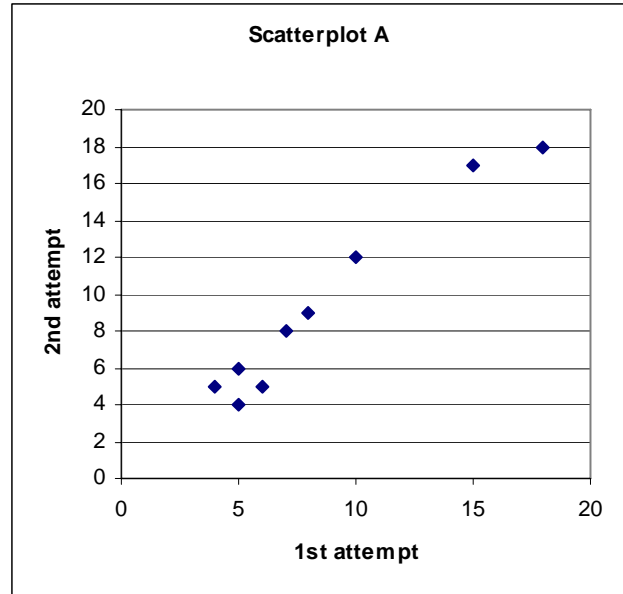
How might we gather data to test our predictions?

Transparency 2

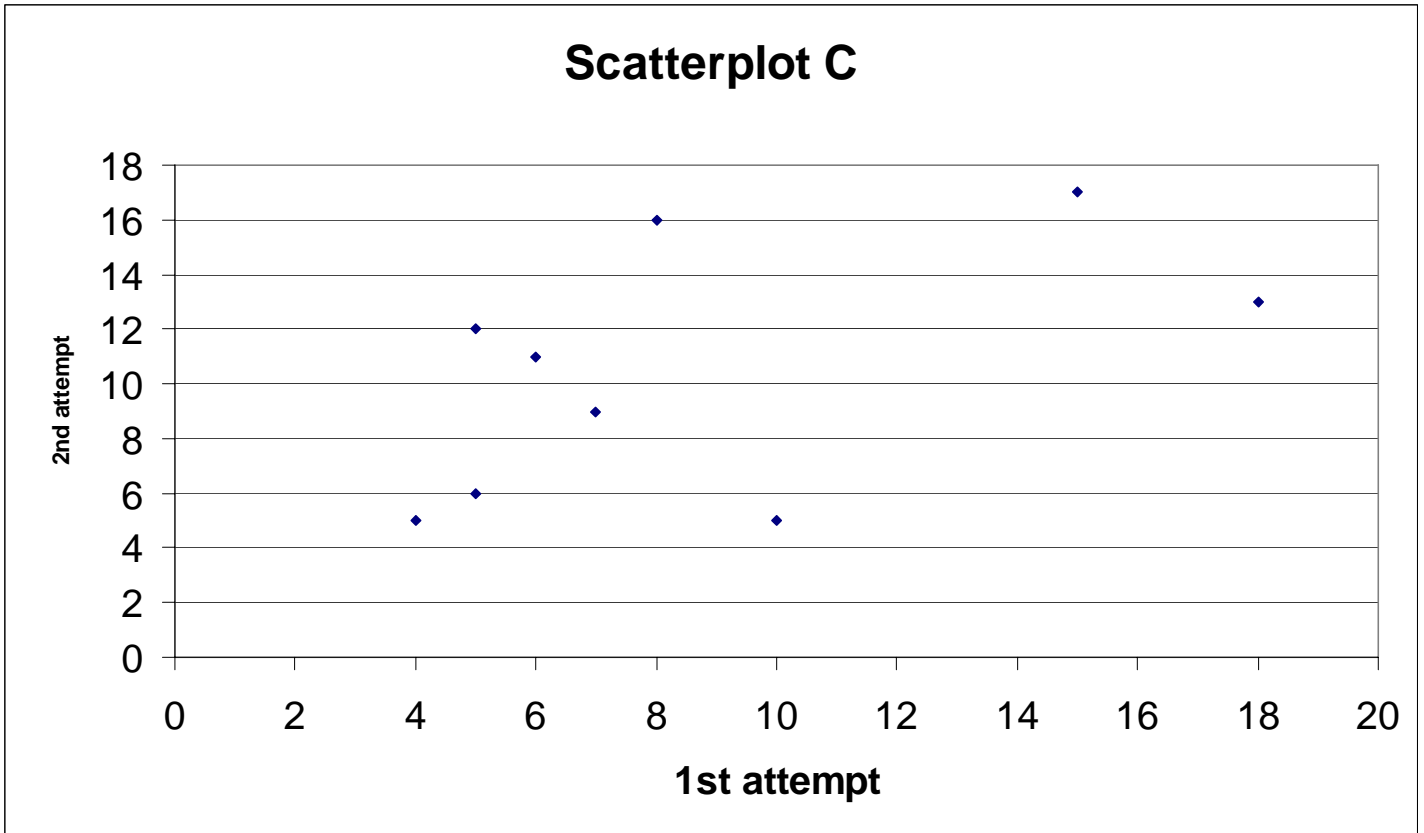
Student	# Revolutions on 1 st Attempt	# Revolutions on 2 nd Attempt

Transparency 3

1. If possible, sketch a trendline.
2. Predict the number of revolutions on the 2nd attempt if the number on the 1st attempt was 13...30...100.

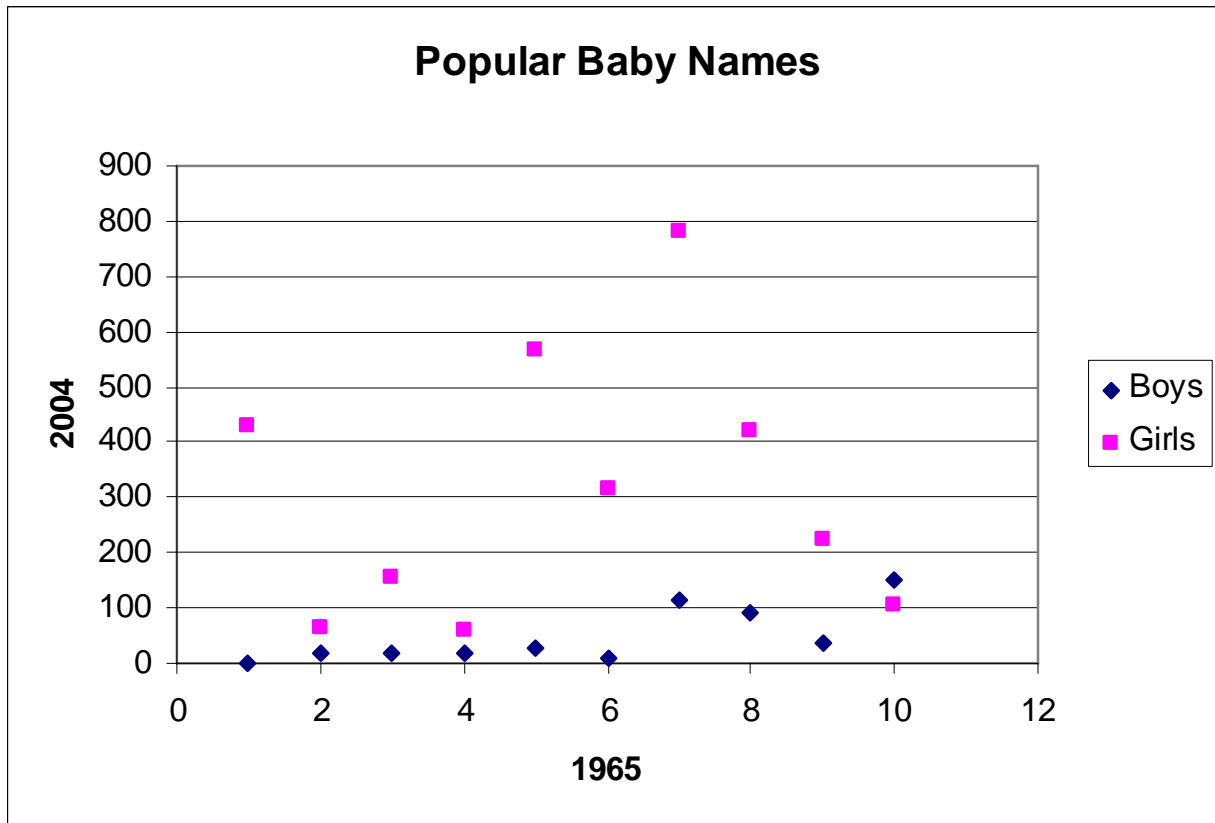


Transparency 4



	1 st attempt	2 nd attempt
mean	8.6667	10.44
median	9.0741	10.94

Transparency 5



	1965	2004 – Boys	2004 – Girls
	— — —	- - - - -
Mean	5.5	48.1	312.1
Median	5.5	23.5	269.5
Range	9	147	504

Student Name(s) _____ Date _____

Round and Round

A. Input your class data from Transparency 2. (Use the table that starts in row 6.)

Sketch the resulting scatterplot.

B. For each statement, choose the scatterplot(s) that best represents the situation.

_____ 1. After the 1st attempt, most students were able to increase the number of revolutions on their 2nd attempt.

_____ 2. After the 1st attempt, most students made fewer revolutions on their 2nd attempt.

_____ 3. The number of revolutions on the 1st attempt is about the same as the number of revolutions on the 2nd attempt.

_____ 4. There is not a strong relationship between the number of revolutions made in the two attempts.

_____ 5. Most students did considerably better on their 2nd attempt than on their 1st attempt.

_____ 6. Based on the data you have from your class, which scatterplot would look most like yours? Explain.

Student Name(s) _____ Date _____

Round and Round (continued)

7. Now that you have analyzed possible scenarios for scatterplots A, B, and C, write a statement that describes the relationship between the 1st attempt and 2nd attempt for your class.

8. Use formulas to calculate the mean, median, and mode of the data for the 1st attempt and for the 2nd attempt. Record the results below.

	1st attempt	2nd attempt
mean		
median		
mode		

9. Which measure of central tendency best describes the number of revolutions made on the 1st attempt and 2nd attempt? Explain your choice.

Baby Names

(based on data from Social Security card applications)

Given: Some baby names are more popular (occur more often) than others. The list of the most popular baby names changes from year to year. While some names are used less over time, others remain popular.

Question: Over the last 40 years, do you think boy names or girl names have been less "trendy"? In other words, do you think children in your generation are more likely to have the same names as adults in your parents' generation if they are boys or girls? Today you will research to compare the ranking of the most popular names in the year 1965 to the ranking of those names in the year 2004 (a span of 40 years).

Directions:

1. Go to the website below to determine the top 10 names for boys and girls in the year 1965.
2. Record the names missing in the tables below.
3. Look up ranks missing for each name for the year 2004 and add that data to the chart. (The database lists the top 1000 names. Use a rank of 1001 if a name is not included.)

<http://www.ssa.gov/OACT/babynames/>

BOYS		
Name	1965 Rank	2004 Rank
	1	
John	2	18
David	3	
James	4	17
	5	
William	6	
	7	113
Richard	8	
Thomas	9	37
Jeffrey	10	

GIRLS		
Name	1965 Rank	2004 Rank
	1	
Mary	2	
	3	154
Kimberly	4	
Susan	5	565
Patricia	6	
Donna	7	781
	8	
Cynthia	9	
Angela	10	105

Baby Names

4. Looking at the data in the table, what do you notice about the change in rank of the top 10 names for boys versus the change in rank of the top 10 names for girls over the last 40 years?

5. Looking at the data in the scatterplot, explain how any observations you made from the table in #4 are reflected in the scatterplot.

6. Calculate the mean rank of the given boy names for 2004 and girl names for 2004. How many boy names and how many girl names were more popular than the mean rank?

7. Calculate the median rank of the given boy names for 2004 and girl names for 2004. How many boy names and how many girl names were more popular than the median rank?

8. Calculate the range in the rankings of the given boy names for 2004 and girl names for 2004. How does the range value for the boys compare to that of the girls? What does this mean?

9. Determine whether the mean or median best describes the data. Position the lines below the scatterplot (solid for boys and dashed for girls) on the scatterplot to represent these measurements. What does the position of the lines on the scatterplot emphasize about the relationship between how the popularity of the top ten names for boys and girls has changed from your parents' generation to now?

10. Describe the relationship between the points on the scatterplot and the lines you drew in #9 for the boy names and the girl names. What does this mean?

What's In A Name?

1. Access the website <http://www.ssa.gov/OACT/babynames/> . In an earlier activity we compared the ranking of the top ten names of your parents' generation (1965) to the ranking of those names today to answer the question about how the popularity of names stands the test of time.
 - a. Consider the following set of questions.
How has the number of people having the most popular boy name changed over the last 10 years? How many people do you predict might have the most popular name in 2010?

- b. Fill in the table using the website.

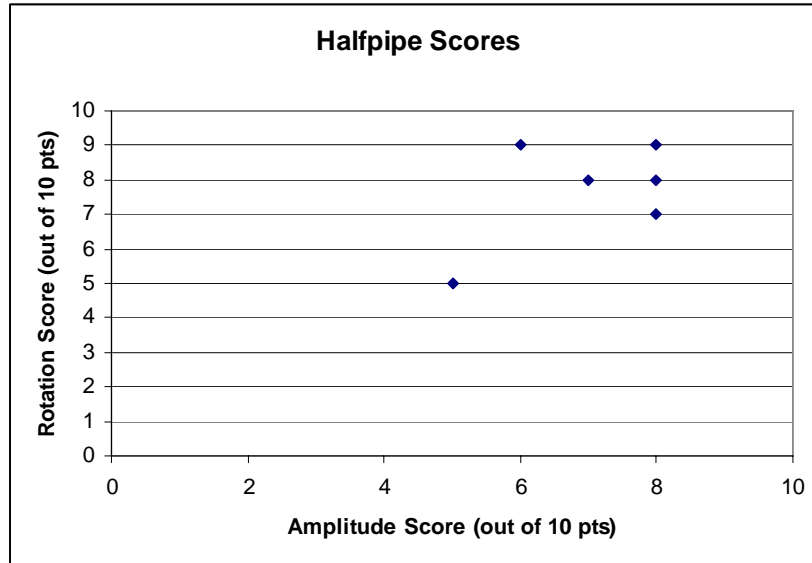
?	?

- c. Use the TI-73 to create a scatterplot.
 - d. Draw a trendline if appropriate.
 - e. Calculate the mean, median, and range of your data.

Minimum	
Maximum	
Mean (average)	
Median	
Mode	
Range	

- f. Respond to the questions in part a. Justify your answers using the scatterplot, trendline, and/or statistical measurements to support your conclusions.

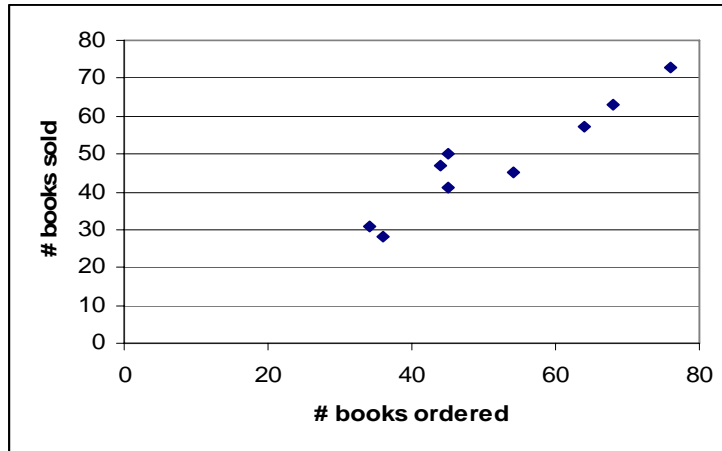
1. The scatterplot below compares the score for amplitude (height) to the score for rotations (spins and flips) for six skateboarders at the weekend meet.



- Which of the following statements would be supported by the scatterplot?
- As the score for amplitude increases, the score for rotations tends to increase.
 - As the score for amplitude increases, the score for rotations tends to decrease.
 - As the score for rotations increases, the score for amplitude tends to decrease.
 - The score for rotations tends to be the same as the score for amplitude.

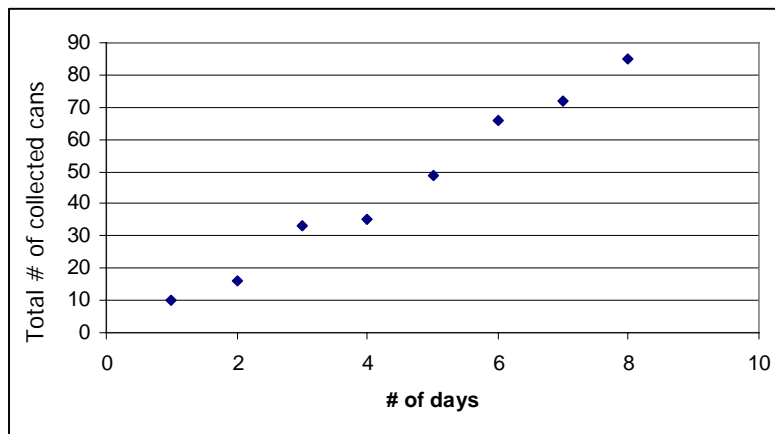
2. Which relationship, when graphed on a scatterplot, would **not** be described as having a positive trend?
- Height of a student compared to his/her weight
 - The amount of money earned babysitting compared to the number of hours spent babysitting
 - The number of miles driven compared to the amount of gas in the tank of the car
 - All of the above relationships have a positive trend.

3. The following scatterplot compares the number of books ordered through the school fund raiser to the number of books that were actually paid for and sold.



If the mean(average) number of books ordered is about 52, estimate the mean(average) number of books sold based on the trends in data in the scatterplot.

- A. greater than 49
 - B. between 47 and 49
 - C. between 45 and 47
 - D. less than 45
4. Ms. Smith's class is collecting aluminum cans for a recycling project as shown in the scatterplot below.



At this rate, about how many days will it take to collect 150 cans?

- A. 15 days
- B. 10 days
- C. 20 days
- D. 150 days