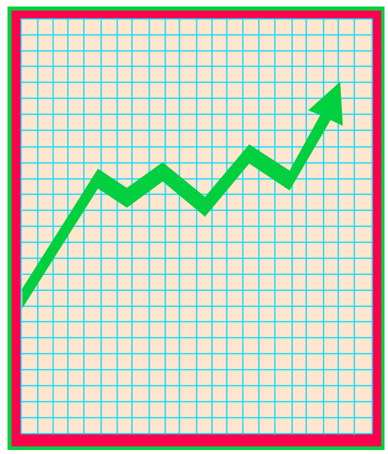
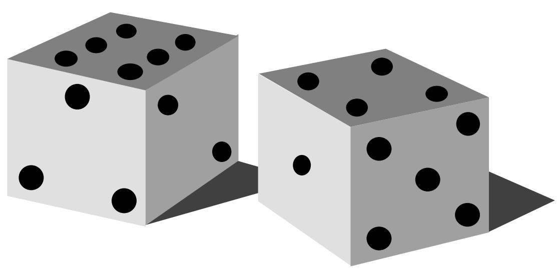
Maryland State Mathematics
Core Learning Goal 3:

Data Analysis and Probability

A Collection of Student Activities





Terry Cepaitis
Anne Arundel County

Mary Jo Messenger
River Hill High School, Clarksville, Maryland

Maryland State Mathematics
Core Learning Goal 3:
Data Analysis and Probability
A Collection of Student Activities

The purpose of these materials is to provide resources for teachers as they prepare their students for Maryland Mathematics High School Assessment I. Non-Commercial reproduction and use of these materials is permitted and encouraged.

Additional items may be added to this collection in the future. This version was completed August 12, 2000; the latest version of this booklet may be downloaded in PDF format from http://www.messengerconnection.com **under "Teacher Resources" on the menu.**

Terry Cepaitis
Anne Arundel County

Mary Jo Messenger
River Hill High School, Clarksville, Maryland

Table of Contents

Core Learning Goal 3: Data Analysis and Probability
A Collection of Student Activities

Unit Outline

The following Unit Outline is offered as a first attempt to identify good sources of activities and background information for teachers who are preparing students to pass the Maryland State High School Assessment 1 (HSA 1) on Algebra and Data Analysis. The sources emphasize active student participation and the use of real data.

Core Learning Goal 3: Data Analysis and Probability

The student will demonstrate the ability to apply probability and statistical methods for representing and interpreting and communicating results, using technology when needed.

I. Measures of Central Tendency and Variability The students will use measures of central tendency and variability to solve problems and make informed decisions. (CLG 3.1.2)	
Central tendencies Mean Median Mode	• *Exploring Data* A8, A9 • *Contemporary Mathematics in Context, Course 1, Part A,* Unit 1, Lesson 2 Shapes and Centers • *Mathematics in a World of Data*, Lesson 6 • *Exploring Centers*, Lessons 2, 3, 4
Variability Range Interquartile range Quartile Solve problems Make informed decisions	• *Exploring Data* A10, A11, A12, A13, A17 • *Contemporary Mathematics in Context, Course 1, Part A,* Unit 1, Lesson 3 Variability • *Mathematics in a World of Data*, Lesson 12, 13
II. Proper and Improper Use of Statistics Given a set of data or statistics, the students will analyze and identify both proper and improper use of statistics. (CLG 3.2.3)	
Communicate the use and misuse of statistics Misuse of scaling Inappropriate measure of central tendency Misuse of 3D figures Data bias Predicating outside the domain	• *The Visual Display of Quantitative Information*, Chapter 2, Graphical Integrity • *The Visual Display of Quantitative Information*, Chapter 3, Sources of Graphical Integrity • *Chance Website* Chance News • *How to Lie with Statistics*

III. Probability and Simulations Using data, the students will determine the experimental or theoretical probability of an event. (CLG 3.1.3)	
Theoretical probability	• *Exploring Probability* A14, A15, A16, A17, A18, A19, A23, A, 24, A25, A26, A30, A31, A32, A33 • *Mathematics in a World of Data*, Lesson 11
Use simulations	• *The Art and Technique of Simulation* A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, A11, A12, A13, A14, A15, A16, A17 • *Contemporary Mathematics in Context, Course 1, Part B*, Unit 7, Lesson 1 Simulating Chance Situations Unit 7, Lesson 2 Estimating Expected Values and Probabilities Unit 7, Lesson 3 Simulation and the Law of Large Numbers • *Probability Through Data*, Lessons 3 and 4
Use statistical inferences	• *The Art and Technique of Simulation* A12, A13
Estimate probability	• *The Art and Technique of Simulation* A6, A7, A8, A9, A10, A11
Given data from simulation or research, the students will make informed decisions and predictions. (CLG 3.2.1)	
Based on data from simulations or research Make informed decisions Make predictions	• *Exploring Surveys and Information from Samples* A27, A28, A19 A2, A3, A4, A5, A6 • *Mathematics in a World of Data*, Lesson 9

IV. Experimental Design The students will describe how they would do an investigation, select an investigation, and defend their choice. Students will consider simple random sampling (SRS) techniques that may include sampling size, bias representation, and randomness. (CLG 3.1.1)	
Simple Random Sample Sample size Biased representation Describe how to do an investigation Select an investigation Defend their choice Analyze data	• *Exploring Surveys and Information from Samples* A17, A18, A19, A20 • *Activity-Based Statistics* – "Random Rectangles" • *Mathematics in a World of Data*, Lesson 5 • *Probability through Data*, Lesson 5
V. Line of Best Fit The students will demonstrate their understanding of the process by finding a line of best fit and by using it to make predictions and/or interpret data (slope and intercepts) or by using a curve of best fit to make a prediction. (CLG 3.2.2)	
Line of best fit Interpret data Make predictions Given curve of best fit Interpret data Make predications Interpolate/extrapolate	• *Exploring Data* A20, A 21, A22, A23, A24, A25, A26, A27, A28, , A29, A30, A31, A32, A33, A34, • *Exploring Regression*, Lessons 1-6 • *Contemporary Mathematics in Context*, *Course 2*, *Part A*, Unit 3, Patterns of Association

Core Learning Goal 3: Data Analysis and Probability
A Collection of Student Activities

Notes for Teachers

The following notes are offered as a brief overview of the concepts that are covered in the activities presented in this collection. Also included are comments emphasizing the importance of classroom process and the use of technology.

During the past 15 years there has been a significant increase in the need for elementary and secondary students to study statistics and a significant increase in the materials available to help with that endeavor. Teachers are encouraged to seek out these resources and to use the print materials cited in the Unit Outline and listed at the end of this collection. All these activities are designed to help students become better thinkers, and better citizens.

Measures of Central Tendency and Variability

Students enter high school with a good understanding of how to determine the mean, median, and mode. What they need at this point is to understand why one measure of central tendency might be preferable to another in certain situations. The mean is probably overused, and the median in many cases would probably be a better representative measure. For example, if Bill Gates were a member of your high school graduating class and if the average salary of the class were given as the mean, the figure would be misleading. The median is an appropriate measure of the center for data sets such as this one where a plot of the data would be skewed. In general, a plot is very helpful in determining which measure of central tendency is best to use. If the plot is mounded and symmetrical, the mean and the median are approximately the same. In this case the mean is often used. If the plot is skewed, however, then the median is probably the best measure of center to use. Activities that help students understand these ideas include: *The Effect of Outliers on the Mean and Median of a Data Set*, and *Measures of Central Tendency and a "Typical" Value for a Data Set*.

The measure of center does not tell all that is essential to understanding a data set. Students also need to understand how data varies around the center. Is it tightly clustered, or is it spread out over a wide range? These are questions that students need to understand as they explore measures of variability. For the Maryland State High School Math Assessment 1, students need to understand and to be able to calculate the range and the interquartile range for a data set. Students often are confused by the idea of "range," thinking that it means, for example, from 10 to 40. The range and the interquartile range are both numbers. If the numbers in a data set vary from 10 to 40, then the range would be 30. Activities that cover these topics include: *Measures of Central Tendency and Five Number Summaries*, *The Five Number Summary and the Graphing Calculator*, *The Five Number Summary and Measures of Variability*, *Displays of Variability and Boxplots*, *Boxplots and the Graphing Calculator*, *Boxplots and Measures of Central Tendency*, *Problem Solving and Informed Decisions Using Statistical Measures*, and *Outliers for Data Sets*.

Proper and Improper Use of Statistics

This is a very important topic in the study of statistics for two reasons. First, it requires students to analyze data and statistics often in graphical form to determine whether it is valid. Second, it helps students realize that they always need to evaluate whatever they read or see from a critical point of view. Students need to be trained to be thinking consumers and citizens.

Edward Tufte has two excellent chapters on graphical integrity in his book, *The Visual Display of Quantitative Information*. In this book, he cites six principles that graphical displays should follow:

- Representation of numbers on a graph should be directly proportional to the numerical quantities represented.
- Graphs should be very clearly labeled and explanations written out on the graphic itself.
- Graphs should emphasize the data variation and not the graphical design.
- Graphical displays involving money over time should use deflated and standardized units.
- The number of variable dimensions shown in the graph should not be greater than the number of dimensions in the data.
- Graphs must not give data out of context.

When these principles are not followed, graphical displays are often misleading.

Along with the misuse of graphical displays, students also need to understand the importance of data bias and of not predicting outside a reasonable domain. The activities in this section explore these ideas.

Probability and Simulations

For many situations, specific outcomes cannot be predicted with certainty. Will the plane arrive on time? Will my luggage be lost? Will you be late to school tomorrow? Over the long run, however, we can establish how frequently these things happen. This relative frequency over the long run is called the probability of an event—it is the measure of chance or likelihood that the event will occur again. Students need to understand this concept of relative frequency. Students also need to know how to estimate probabilities from real data. The activities *Understanding Probability*, *Tumbling Thumb Tacks*, *Using Results of a Study*, and *Collecting Your Own Data* help students develop these concepts.

What is the difference between experimental probability and theoretical probability? How are these ideas related? Students need to understand this difference. They also need to experience the law of large numbers; that is, the more times an experiment is conducted, the closer the experimental results come to the theoretical results. The activity, *Spinning: From the Experimental to the Theoretical*, will help students with these ideas.

A very important part of this particular section is the concept and application of simulations. Simulation is a modeling technique used to help people answer questions about real situations by using probability experiments that are similar to the real situations. These are the basic steps in the simulation process:

- State the problem.
- State the assumptions.

- Select a model to generate outcomes.
- Simulate many repetitions.
- Analyze the results, and state the conclusion.

Students should conduct many investigations using simulations. Coins, spinners, and number cubes should be used first, followed by calculator or computer random number generators. The calculator and computer have the advantage of speed and ease, but the hands-on experiences are fundamentally important in helping students see that estimating probabilities is a dynamic process with variability in the outcomes.

The first two simulation activities—*Figuring Free-Throws*, and *Thinking About Three*—involve a probability of 0.5. These can easily be modeled using coins or number cubes. The activity, *A Test Simulation*, involves a probability of 0.2 and would work well with a spinner or a calculator. *Figuring Free-Throws Again* involves a probability of 0.67 and either number cubes or a calculator would be easiest.

Another important concept in this section is that of expected values. Students need to know how to find and use expected values. A simple example of such a problem involves determining the number of boxes of cereal you would need to buy on average in order to get all three prizes that a cereal company is offering. Assume that each box of cereal has one prize and the prizes are distributed evenly among the boxes. Simulate this situation by generating numbers from 1 through 3, and count the number of digits you have to generate before you have a 1, a 2, and a 3. Doing this 30 times gave the following results: 7, 3, 5, 5, 5, 7, 4, 7, 8, 4, 4, 3, 4, 4, 5, 4, 7, 6, 4, 14, 3, 3, 5, 3, 6, 4, 7, 4, 5, 3. This list of data shows that for the first trial 7 digits were randomly generated before a 1, a 2, and a 3 appeared. For the next trial, 3 digits were generated showing a 1, a 2, and a 3. The list continues showing the results of 30 trials. The mean of this data is 5.1, or on average we can expect to have to buy approximately 5 boxes of cereal to get all three prizes.

The theoretical average number of boxes purchased is 5.5. In general, the theoretical waiting time to get n equally likely prizes is

$$\frac{n}{n}+\frac{n}{n-1}+\frac{n}{n-2}+\ldots+\frac{n}{3}+\frac{n}{2}+\frac{n}{1}.$$

This formula is based on the fact that if the probability of an event is p, the average or expected number of trials before it occurs is $\frac{1}{p}$. So if we have 1 of the 3 prizes already, then the probability of getting a new prize in the box of cereal is $\frac{2}{3}$. This means that, on average, we would have to purchase $\frac{3}{2}$ additional boxes to get a new prize. Problems dealing with this concept include *Simulation with an Unknown Number of Trials* and *Calling for the Newspaper*.

Experimental Design

Students need to understand the concepts of a simple random sample, sampling size, and biased representation in order to be prepared for the High School Assessment. A sample is random if both these conditions are true: each member of the population is equally likely to be chosen, and the members of the sample are chosen independently of one another. It is often not possible to pick a simple random sample, however, and in such cases bias may be introduced. Bias may be present in results because those results are based on a voluntary response sample.

Bias might also be due to undercoverage, nonresponse, or poorly worded questions. In addition to these ideas, students need to understand that in order to be confident about the results of a survey or an experiment, they must conduct enough trials or they must survey enough randomly selected people to get good results. The larger the number of trials or people surveyed, the more confident students can be in the results. In general, these concepts are not treated in a formal way in these activities, but students need to have a fundamental understanding of these ideas. Activities that emphasize these concepts are *Random Samples*, *Simple Random Samples*, *Random Digit Tables*, and *Random Rectangles*.

Line of Best Fit

Students need to be able to determine a line of best fit and use that line to answer questions or make predictions. They need to be able to interpret the y-intercept and the slope in terms of the context of the problem. They need to totally understand the process and the relationships among the table of values, the symbolic rule, and the graph. In addition to determining the line of best fit, students need to be able to make sense out of a curve of best fit. They should be able to interpret data and make predictions using the table, the equation, or the graph.

The Importance of Process

The National Council of Teachers of Mathematics (NCTM) Principles and Standards and the Maryland State Department of Education Skills for Success emphasize the importance of how we go about doing things in the classroom. It is not enough just to expose students to information. How we conduct our classrooms is more important now than ever before as we learn from employers what is important in our country's workforce. The process standards of NCTM mandate that we address problem solving, reasoning and proof, communication, connections, and representation. In order to accomplish this, we need to have our students work together, investigating and discussing their ideas with their group and with the entire class. We need to allow plenty of time for class discussion. Having students make sense out of what they have experienced or investigated is critically important. Answering questions such as the following is a very important part of what students need to be experiencing in the classroom.

- Do your results agree with what you expected to find?
- Would you expect to get the same results if you repeated this investigation?
- What would happen if you used a larger sample size?
- How do your results compare to the results of the rest of the class?
- What conclusions can you make based on these results?

NCTM Mathematics Process Standards	
Problem Solving	Students should build new mathematical knowledge through problem solving; solve problems that arise in mathematics and in other contexts; apply and adapt a variety of appropriate strategies to solve problems; monitor and reflect on the process of mathematical problem solving.
Reasoning and Proof	Students should recognize reasoning and proof as fundamental aspects of mathematics; make and investigate mathematical conjectures; develop and evaluate mathematical arguments and proofs; and select and use various types of reasoning and methods of proof.
Communication	Students should organize and consolidate their mathematical thinking through communication; communicate their mathematical thinking coherently and clearly to peers, teachers, and others; and analyze and evaluate the mathematical thinking and strategies of others.
Connections	Students should recognize and use connections among mathematical ideas; understand how mathematical ideas interconnect and build on one another to produce a coherent whole; recognize and apply mathematics in contexts outside of mathematics.
Representation	Students should create and use representations to organize, record, and communicate mathematical ideas; select, apply and translate among mathematical representations to solve problems; use representations to model and interpret physical, social, and mathematical phenomena.

The Importance of Technology

In the last few years, the graphing calculator has changed both what we teach in mathematics and how we teach it. Students need to understand the relationships among a graph, the corresponding table of values, and the related equation. Students need to be able to make a scatter plot, evaluate the pattern, calculate a model, and use that model to make predictions. They also need to be able to conduct simulations and understand them as models for answering questions about real situations. The graphing calculator is the powerful tool that enables students to conduct these activities and investigate these concepts most effectively and efficiently. For that reason, calculators should be an integral part of mathematics instruction at all levels, and students need to have access to graphing calculators both in and out of the classroom. If students, especially our most at-risk students, are to be prepared for the high school assessment, they need to use the graphing calculator on a daily basis.

Core Learning Goal 3:
Data Analysis and Probability

I. Measures of Central Tendency and Variability

Indicator 3.1.2

The students will use measures of central tendency and variability to solve problems and make informed decisions.

Notes for Teachers

Students enter high school with a good understanding of how to determine the mean, median, and mode. What they need at this point is to understand why one measure of central tendency might be preferable to another in certain situations. The mean is probably overused, and the median in many cases would probably be a better representative measure. For example, if Bill Gates were a member of your high school graduating class and if the average salary of the class were given as the mean, the figure would be misleading. The median is an appropriate measure of the center for data sets such as this one where a plot of the data would be skewed. In general, a plot is very helpful in determining which measure of central tendency is best to use. If the plot is mounded and symmetrical, the mean and the median are approximately the same. In this case the mean is often used. If the plot is skewed, however, then the median is probably the best measure of center to use. Activities that help students understand these ideas include: *The Effect of Outliers on the Mean and Median of a Data Set*, and *Measures of Central Tendency and a "Typical" Value for a Data Set*.

The measure of center does not tell all that is essential to understanding a data set. Students also need to understand how data varies around the center. Is it tightly clustered, or is it spread out over a wide range? These are questions that students need to understand as they explore measures of variability. For the Maryland State High School Math Assessment 1, students need to understand and to be able to calculate the range and the interquartile range for a data set. Students often are confused by the idea of "range," thinking that it means, for example, from 10 to 40. The range and the interquartile range are both numbers. If the numbers in a data set vary from 10 to 40, then the range would be 30. Activities that cover these topics include: *Measures of Central Tendency and Five Number Summaries*, *The Five Number Summary and the Graphing Calculator*, *The Five Number Summary and Measures of Variability*, *Displays of Variability and Boxplots, Boxplots and the Graphing Calculator*, *Boxplots and Measures of Central Tendency, Problem Solving and Informed Decisions Using Statistical Measures*, and *Outliers for Data Sets*.

The Effect of Outliers on the Mean and Median of a Data Set

1. Calculate the mean and median for each of the following data sets.

    a. -6 -4 -2 2 4 6

    b. 6 6 6 6 6

    c. 25 25 35 35 40 45 45 55 55

2. What do you notice about the relationship between the mean and the median for each of the data sets in #1?

3. What characteristic of each data set produces this relationship?

4. Generate a data set of five values that has the same mean and median.

Consider the following data set: 6 7 8 9 10

5. Without using a calculator, what are the mean and median of this data set?

6. Suppose we change the data set to: 6 7 8 9 **100**

    a. Compute the mean and the median of the new data set.

    b. Describe the relationship between the mean and the median.

    c. What feature of the new data set yields this relationship?

7. Suppose we change the data set to: **0** 7 8 9 10

    a. Compute the mean and the median of the new data set.

    b. Describe the relationship between the mean and the median.

    c. What feature of the new data set yields this relationship?

An *outlier* is a value in a set of data that is significantly greater than or less than the other elements in the set.

8. What effect did the outlier have on the relationship between the mean and median in #6?

9. What effect did the outlier have on the relationship between the mean and median in #7?

10. Generate a set of five values that has a mean that is less than the median.

11. Generate a set of five values that has a mean that is greater than the median.

12. Given the following data set:

    0 49 49 49 50 51 51 51 100

    a. Does this set of data have any outliers?

    b. What are the mean and median of this set of data?

    c. What effect do the outliers have on the relationship between the mean and the median of this data set?

13. What effect do outliers have on the relationship between the mean and the median of a data set?

14. Examine the following data sets. Without using your calculator, describe the relationship between the mean and the median.

    a. 5 8 15 23 25 78

    b. -6 -6 -3 2 7 10 10

    c. -3 3 3 3 3

    d. 1 499 500 501 999

Measures of Central Tendency and a "Typical" Value for a Data Set

The mean is affected by outliers and the median is not, so we say that the *median is resistant to outliers*. For this reason, when a data set contains outliers, it is usually more appropriate to use the median to describe a "*typical*" value in the set.

1. The table shows population data from the 1990 Census for several South American countries. If you were reporting these data in a magazine article, which measure of central tendency would you use to describe a "typical" value for the population of these countries? Explain.

Country	Population in Millions
Argentina	32
Bolivia	7
Brazil	150
Colombia	33
Ecuador	10
Paraguay	5
Peru	22
Uruguay	3
Venezuela	19

2. After playing several concerts around the country, the band *Jive Talkin'* reported that their tour was a huge success. The lead singer told reporters that the average attendance at the band's concerts was 25,000! The attendance at each of the four concerts is shown in the table.

City	Attendance
Albany	500
Dallas	300
Phoenix	99,000
Washington	200

   a. Were the lead singer's calculations correct?
   b. Explain why the claim is misleading.
   c. Which measure of central tendency would be a more appropriate reflection of the band's typical concert attendance?

3. A soft drink company asked 8 volunteers to taste their new diet soda and rate it on a scale of 1 to 10 (10 being the best score). The results are:

   1 1 1 1 1 1 1 10

   If you were a salesperson for the company, which measure of central tendency would you report to your prospective clients? Explain.

Transformations and Measures of Central Tendency

The results of a math test for 30 students are as follows:

89	67	99	92	66	52	74	74	85	45
77	93	96	82	75	60	55	90	81	90
27	78	86	100	92	68	75	81	74	57

1. Enter the test scores into list L1 on your calculator.

2. Determine the following measures of central tendency:

   mean: median: mode:

3. Suppose the teacher made an error in calculating the test scores, and the actual scores should be 5 points *less* than those originally reported. In your list window, subtract 5 from each of the test scores in your list by placing the cursor on top of list L1 and typing L1-5. Press [ENTER].

4. Recalculate the mean, median, and mode for the new test scores.

   mean: median: mode:

5. Compare the mean, median, and mode of the new test scores to those of the original scores. Describe this relationship.

6. Suppose the teacher added two points to each of the reduced test scores. Without using your calculator, what are the mean, median, and mode?

7. What happens to the measures of central tendency if a constant is added to each value in a data set?

8. What happens to the measures of central tendency if a constant is subtracted from each value in a data set?

9. The mean age of six students is 14. What will their mean age be in five years?

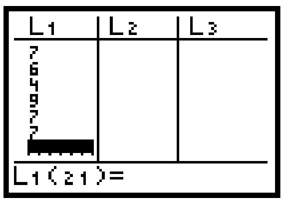
10. The median age of six students is 14. What was their median age five years ago?

Calculator Instructions:
Entering Data and Measures of Central Tendency

1. Clear lists of unwanted entries.

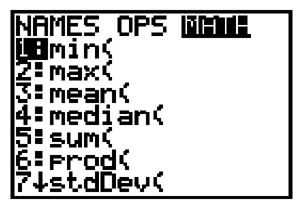
   Press [2nd][STAT]. Use the arrow key to place the cursor on top of L_1.
   Press [CLEAR][ENTER]. Use the arrow key to place the cursor in L_1.

2. Enter data into list. Enter each numerical value followed by [ENTER]. As you type, the value is displayed on the bottom line. When you have finished, your display should look similar to this:



3. Calculate mean and median.

   Press [2nd][MODE] to return to the home screen.
   Press [2nd][STAT] and arrow over to MATH to access the List Math menu.
   Your screen should look like the display shown.



   · Calculate the mean of the data.
     Press [3] to select the mean option.
     Enter the list name for the data. Press [2nd][1].
     Press [ENTER].
     The mean should be displayed on the screen.

   · Calculate the median of the data.
     Press [4] to select the median option.
     Enter the list name for the data. Press [2nd][1].
     Press [ENTER].
     The median should be displayed on the screen.

4. Calculate the mode of the data.

   Press [STAT] and choose SortA(.
   Enter the list name for the data. Press [2nd][1].
   Press [ENTER]. This will sort the data from least to greatest.
   Look at the sorted list. Press [STAT] and choose EDIT.
   Choose the value that occurs most often.

Measures of Central Tendency and Frequency Tables

The grades on a test are issued on a four-point scale, where an A is worth 4 points, a B is 3 points, and so on. The grades of students in a math class are shown in the given frequency table.

Grade	Number of Points	Number of Students
A	4	4
B	3	5
C	2	12
D	1	8
E	0	3

Recall that the *mean* of a set of values is $\dfrac{\textbf{\textit{total of values}}}{\textbf{\textit{number of values}}}$.

To determine the *mean* test score for the class you need to determine the total points earned and the total number of students in the class.

1. a. How many points are earned by all students with an A?
   How did you determine this amount?
   b. How many points are earned by all students with a B?
   How did you determine this amount?
   c. How many points are earned by all students with a C?
   How did you determine this amount?
   d. How many points are earned by all students with a D?
   How did you determine this amount?
   e. How many points are earned by all students with an E?
   How did you determine this amount?

2. a. What is the total number of points earned by *all* students?
   b. What is the total number of students in the class?
   c. What is the *mean* number of points earned for the class?

3. Recall that the *median* of an ordered set of values is the middle value. What is the *median* number of points earned for the class?
   How did you determine this amount?

4. Recall that the *mode* is the value or values that occur most frequently.
   What is the *mode* number of points earned for the class?
   How did you determine this amount?

Measures of Central Tendency and Five Number Summaries

The table below shows the average monthly high temperatures in Columbus, Ohio, and San Francisco, California.

Month	Columbus, OH	San Francisco, CA
Jan	35	56
Feb	38	59
Mar	49	60
Apr	62	61
May	73	63
Jun	81	64
Jul	84	64
Aug	83	65
Sep	77	69
Oct	65	68
Nov	51	63
Dec	39	57

Source: U.S. National Oceanic and Atmospheric Administration

1. Calculate the mean and median high temperature for both cities.

	Columbus, OH	San Francisco, CA
Mean		
Median		

2. If you looked only at these measures of central tendency when comparing the temperatures in the two cities, what would you conclude?

3. Do you think that the mean and median give a complete description of the temperatures in these cities?

To obtain a better summary of these data, we will identify the values which separate the data into four equal parts. These values are called **quartiles.**

35 38 39 ↑ 49 51 62 ↑ 65 73 77 ↑ 81 83
84

$$Q1 = \frac{39 + 49}{2} = 44 \qquad Q2 = \frac{62 + 65}{2} = 63.5 \qquad Q3 = \frac{77 + 81}{2} = 79$$

The three quartiles are labeled Q1, Q2, and Q3, and represent the **first** or **lower quartile,** the **second quartile** or **median,** and the **third** or **upper quartile**, respectively. These values along with the **maximum** and **minimum** value in the data set make up the **five number summary** of a data set.

The Five Number Summary and the Graphing Calculator

1. What is the five number summary for the average monthly high temperatures in Columbus, Ohio?

Minimum, min	Lower Quartile, Q1	Median, med	Upper Quartile, Q3	Maximum, max

2. a. 75% of the average high temperatures in Columbus are at or above what temperature?
   b. Half of the average high temperatures in Columbus are at or below what temperature?
   c. 25% of the average high temperatures in Columbus are above what temperature?

3. Calculate the five number summary for the average monthly high temperatures in San Francisco, California using a graphing calculator.
   a. Enter the average monthly high temperatures into a calculator list.
   b. To compute the five-number summary:
      - Press [STAT] then arrow over to CALC.
      - Select [1]:1-Var Stats.
      - Select the list containing the appropriate data.
      - Press [ENTER].

      The first value, $\bar{x}$, represents the mean of the data set.
      The sixth value, *n*, gives the number of elements in the set.
      Use the down arrows to access the five-number summary.
   c. What is the five-number summary for the average monthly high temperatures in San Francisco, California?

Minimum, min	Lower Quartile, Q1	Median, med	Upper Quartile, Q3	Maximum, max

4. a. 75% of the average high temperatures in San Francisco are at or above what temperature?
   b. Half of the average high temperatures in San Francisco are at or below what temperature?
   c. 25% of the average high temperatures in San Francisco are above what temperature?

The Five Number Summary and Measures of Variability

The measures of central tendency do not always give a complete numeric description of a set of data. The variation in the data should also be considered. Two measures of variability for a data set are the **range** and the **interquartile range**. The **range** is the difference between the maximum and minimum values. The **interquartile range** is the difference between the upper and lower quartiles.

1. a. Calculate the range and interquartile range for the average high temperatures in Columbus.
   b. What percent of the data is contained in the range?
   c. What percent of the data is contained in the interquartile range?

2. a. Calculate the range and interquartile range for the average high temperatures in San Francisco, CA.
   b. What percent of the data is contained in the range?
   c. What percent of the data is contained in the interquartile range?

3. The measures of central tendency for the average monthly high temperatures in Columbus, Ohio and San Francisco, California are nearly the same.
   a. Compare the ranges and interquartile ranges for the temperatures in these two cities.
   b. How does knowing the measures of variability along with the measures of central tendency enable you to make a better comparison between the average high temperatures in Columbus and San Francisco?

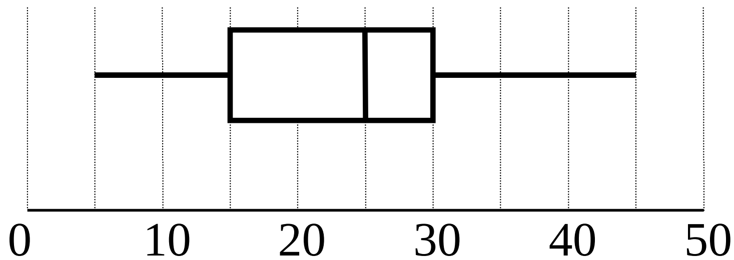
4. Two new sodas were tested and rated on a ten-point scale with 10 being the highest mark. The results are shown below.

	Soda P	*Soda Q*
mean	7	6.7
median	7	7
range	8	4
interquartile range	4	4

   a. Which soft drink had more consistent ratings? Justify your answer using one or two of the above measures.
   b. Which soft drink had the better overall rating? Justify your answer using one or two of the above measures.

Displays of Variability and Boxplots

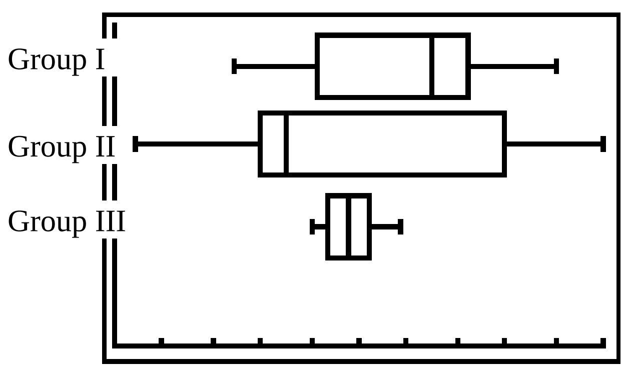
A boxplot an be used to display the variability of a data set. The five number summary is used to draw a boxplot. The display below is a boxplot for a set of data.



The boxplot consists of two segments and a box.. The outer endpoints of the segments are the maximum and minimum of the data. The outer endpoints of the box are the lower and upper quartiles of the data. The line segment inside the box is the median of the data.

1. Identify each value in the five number summary for the given boxplot.
   a. minimum
   b. lower quartile
   c. median
   d. upper quartile
   e. maximum

2. Suppose the box plots on the screen below represent the ages of three different groups of people. Each group consists of the same number of people.



   a. What percent of the ages are at or above the median in each group?
   b. How are the measures of variability represented in the boxplots?
   c. Which group of people were most similar in age?
   d. Which group of people were least similar in age?
   e. Which group had the most young people?

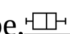
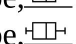
Boxplots and the Graphing Calculator

The table below shows the average monthly high temperatures in Columbus, Ohio and San Francisco, California.

Month	Columbus, OH	San Francisco, CA
Jan	35	56
Feb	38	59
Mar	49	60
Apr	62	61
May	73	63
Jun	81	64
Jul	84	64
Aug	83	65
Sep	77	69
Oct	65	68
Nov	51	63
Dec	39	57

Construct a boxplot for the average monthly high temperatures in Columbus, Ohio and San Francisco, California on a graphing calculator using the following procedure.

1. Enter the Columbus data into list L1 on your calculator.
   Enter the San Francisco data into list L2 on your calculator.

2. Set up the boxplots.
   Clear any equations, press [Y=] and clear all equations.
   Press [2nd][Y=], choose Plot 1, On, Type, 
   Press [2nd][Y=], choose Plot 2, On, Type, 

3. Set up the window.
   Press [ZOOM][9].

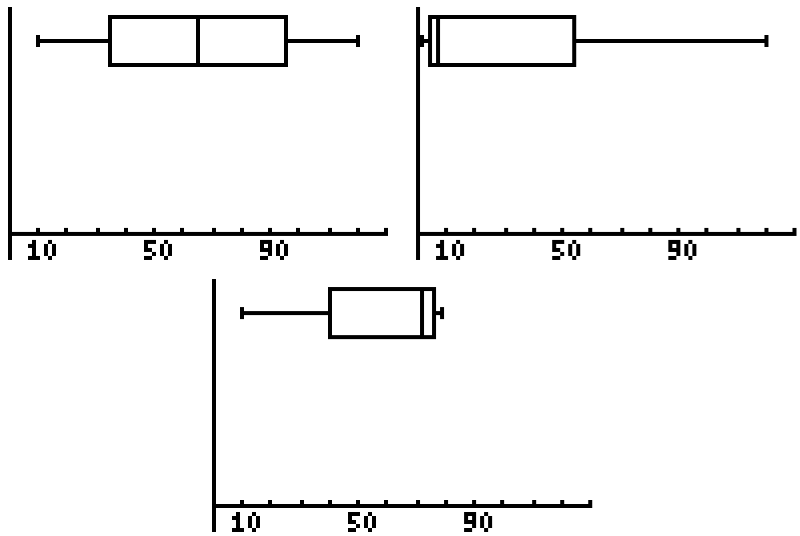
4. Determine the five-number summary by using trace feature.
   a. minimum
   b. lower quartile
   c. median
   d. upper quartile
   e. maximum

5. What do the boxplots tell you about the variability of the average monthly high temperatures in the two cities?

Boxplots and Measures of Central Tendency

The boxplots for three data sets are shown below.

Data Set I **Data Set II** **Data Set III**



1. Compare the shapes of the boxplots.
2. What do these shapes tell you about the variation in the data sets.
3. The data sets for these boxplots are given below.

Data Set I:										
10	20	30	40	50	60	70	80	90	100	110
120										

Data Set II:										
1	2	3	4	5	6	7	8	9	100	110
120										

Data Set III:										
10	10	11	70	71	72	73	74	75	76	78
79										

   a. Calculate the mean of each data set.

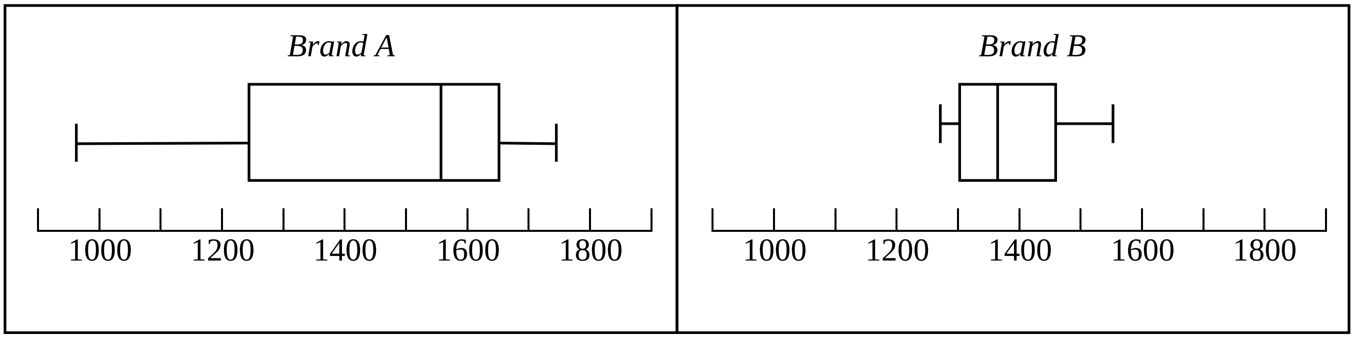
   b. Locate the mean of each data set on its boxplot by drawing a vertical that passes through the mean.

   c. How do the mean and median of each data set compare?

4. What does the shape of a boxplot tell you about the relationship between the mean and the median of its data set?

Problem Solving and Informed Decisions Using Statistical Measures

A manufacturer wants to compare the lifetime of two brands of batteries. Fifty batteries of each type are tested to determine how long each will last. The following boxplots show the number of hours the batteries lasted.



1. Identify the extremes, quartiles, and median for each type of battery.

	Brand A	*Brand B*
minimum		
lower quartile (Q1)		
median		
upper quartile (Q3)		
maximum		

2. How many batteries lasted longer than the median for each brand?

3. If you owned the manufacturing company which measure of central tendency would you advertise as the typical lifetime of a Brand A battery? Justify your choice.

4. Compare the variability in lifetime for the two brands.

   a. How do the ranges compare?

   b. How do the interquartile ranges compare?

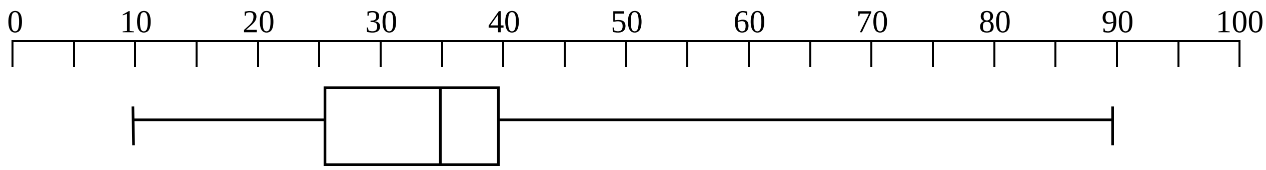
5. Which battery would you want to use in your walkman? Justify your choice.

6. Which battery would you want to use in your smoke detector? Justify your choice.

Outliers for Data Sets

An outlier for a data set is a value that is substantially larger or smaller than the other values in the data set. In this activity you will use the length of the box in a boxplot to determine whether values are outliers. You will need the edge of a sheet of blank paper to mark off 1.5 box lengths.

Look at the boxplot below.



1. Mark off one box length on the edge of the blank sheet of paper. Label the endpoints of this length A and B.

2. Starting at B mark off a second box length on the edge of the paper. Label the endpoint of this second box length C.

3. Locate half of a box length by folding the edge of the paper over so that C is on top of B. Label the crease after making this fold X. The distance between A and X is 1.5 box lengths.

4. To determine the upper outliers, place the edge of your box length paper on the given boxplot so that A lies on the upper quartile and X extends to the right of it. Mark off the distance to X on the boxplot.

   a. What value does this mark correspond to on the number line?

   b. Values that lie above this mark are outliers. Does this data set have any upper outliers?

5. To determine lower outliers, place the edge of your box length paper on the given boxplot so that A lies on the lower quartile and X extends to the left of it. Mark off the distance to X on the boxplot.

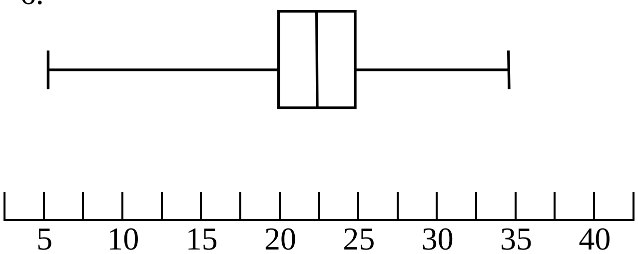
   a. What value does this mark correspond to on the number line?

   b. Values that lie below this mark are outliers. Does this data set have any lower outliers?

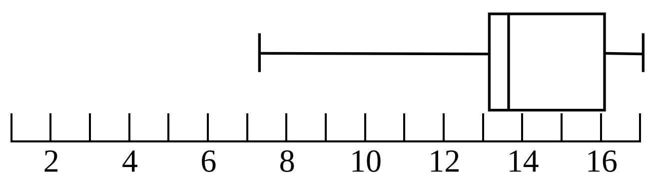
In general for a data set displayed in a boxplot: *values that lie more than 1.5 box lengths above the upper quartile or below the lower quartile are outliers*.

Look at each of the following boxplots. Determine whether or not there are any outliers. If outliers exist, give an interval of values which contain the outliers.

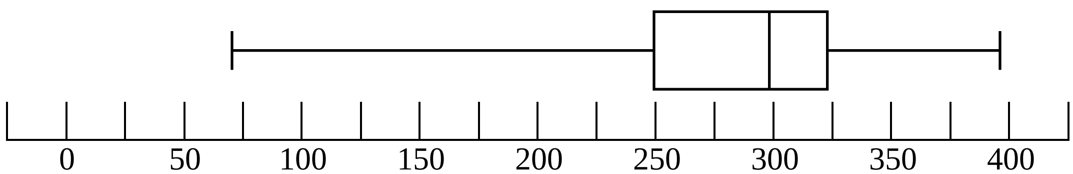
6.



7.



8. Examine the boxplot below.



Tell whether or not each value is an outlier.

a. 50 b. 110 c. 115 d. 375

The table below shows the number of motor vehicle thefts (per 100,000) for twenty states with the highest theft rate.

State	Motor Vehicle Theft/100,000	State	Motor Vehicle Theft/100,000
Alaska	493	Massachusetts	528
Arizona	927	Michigan	701
California	761	Nevada	698
District of Columbia	1,837	New Jersey	581
Florida	721	New Mexico	582
Georgia	629	New York	494
Hawaii	605	Oregon	531
Illinois	490	Tennessee	647
Louisiana	632	Texas	549
Maryland	711	Washington	522

Source: Statistical Abstract of the United States, 1998

9. Construct a boxplot for the data in the table.

10. Are there any outliers in this data set? If so, list the states that are outliers.

Core Learning Goal 3:
Data Analysis and Probability

II. Proper and Improper Use of Statistics

Indicator 3.2.3

Given a set of data or statistics, the student will analyze and identify both proper and improper use of statistics.

Notes for Teachers

This is a very important topic in the study of statistics for two reasons. First, it requires students to analyze data and statistics often in graphical form to determine whether it is valid. Second, it helps students realize that they always need to evaluate whatever they read or see from a critical point of view. Students need to be trained to be thinking consumers and citizens.

Edward Tufte has two excellent chapters on graphical integrity in his book, *The Visual Display of Quantitative Information*. In this book, he cites six principles that graphical displays should follow:

- Representation of numbers on a graph should be directly proportional to the numerical quantities represented.
- Graphs should be very clearly labeled and explanations written out on the graphic itself.
- Graphs should emphasize the data variation and not the graphical design.
- Graphical displays involving money over time should use deflated and standardized units.
- The number of variable dimensions shown in the graph should not be greater than the number of dimensions in the data.
- Graphs must not give data out of context.

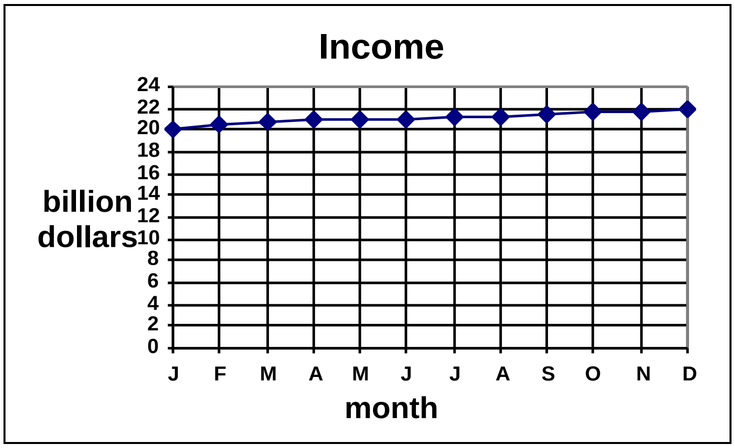
When these principles are not followed, graphical displays are often misleading.

Along with the misuse of graphical displays, students also need to understand the importance of data bias and of not predicting outside a reasonable domain. The activities in this section explore these ideas.

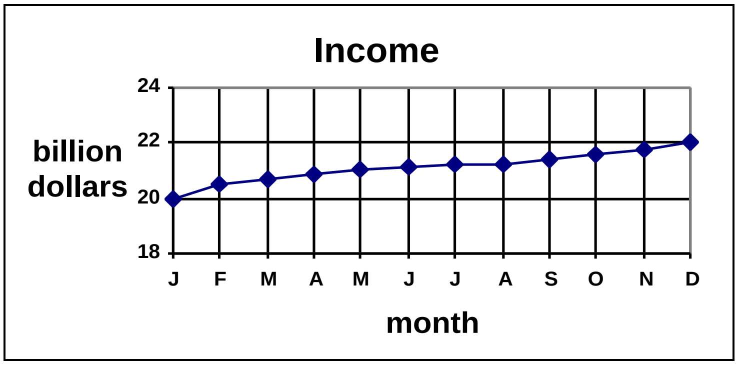
Misuse of Scaling

1. Examine the graphs below. Explain how the different use of scaling affects each graph. Which graph most accurately represents the data?

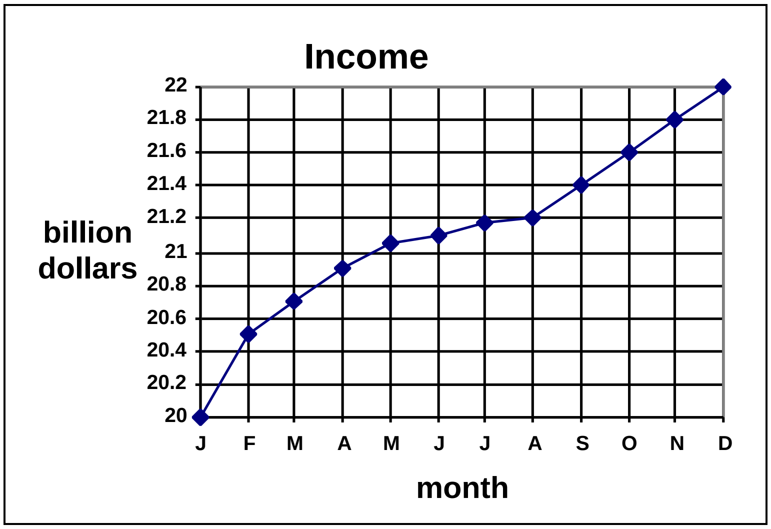
a.



b.



c.



Inappropriate Measure of Central Tendency and Data Bias

From Chance News: Mean statistics: when is the average best?

In an article in the Washington Post, 6 Dec. 1995, by John Schwartz, Schwartz remarks that politicians and others often choose a definition of average that best suits their needs. He tells his readers what mean, median, and mode mean and gives examples of their use and misuse. He starts with the example of John Cannell, who notices that his state's school system claimed high scores on nationally standardized tests and requested test scores from all 50 states. Cannell found that every one claimed to be "above the national average" or the statistical "norm". He called this as the "Wobegan effect". A more detailed discussion of this example can be found in the article *Taking the tests*, Dallas Morning News, 4 Oct. 1994, by Karel Holloway.

As another example, Schwartz remarks that if Bill Gates were to move to a town with 10,000 penniless people the average (mean) income would be more than a million and might suggest that the town is full of millionaires.

Discussion questions:

1. How could the answers Cannell received be correct?

2. Someone once claimed that if any one person moved from state X to state Y, the average intelligence in both states would be increased. Give an explanation of how this could be possible.

Visual Displays of Data

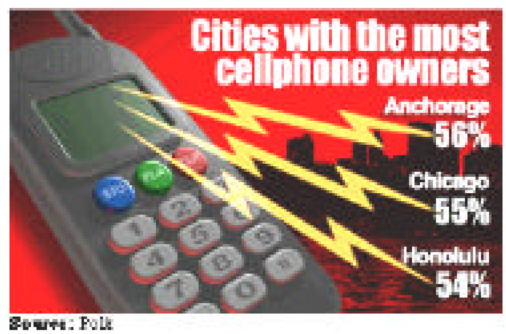
Edward Tufte has two excellent chapters on graphical integrity in his book, *The Visual Display of Quantitative Information*. In this book, he cites six principles that graphical displays should follow:

- Representation of numbers on a graph should be directly proportional to the numerical quantities represented.
- Graphs should be very clearly labeled and explanations written out on the graphic itself.
- Graphs should emphasize the data variation and not the graphical design.
- Graphical displays involving money over time should use deflated and standardized units. The data should be adjusted for inflation.
- The number of variable dimensions shown in the graph should not be greater than the number of dimensions in the data.
- Graphs must not give data out of context.

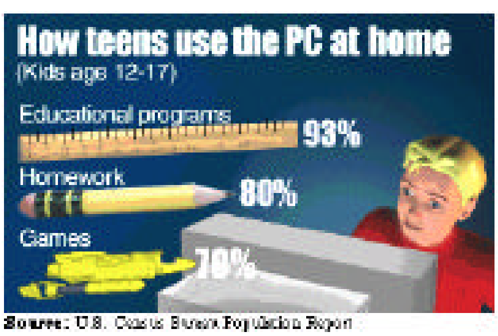
For each of the following graphs write several sentence evaluating the integrity of the graph considering Tufte's principles.

1.

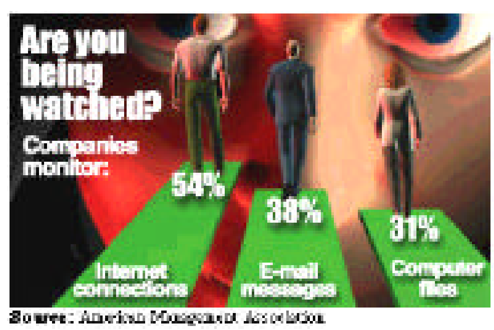
2.



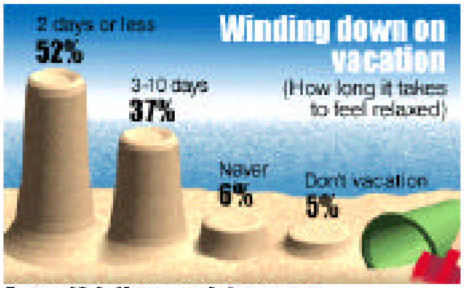
3.



4.



5.



Source: MotherNature.com telephone survey

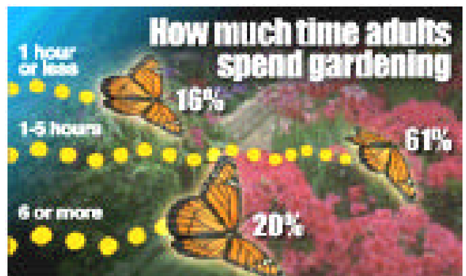
6.

Source: Peter D. Hart Research Associates

7.

Source: IHR Research Group for Dell Computer

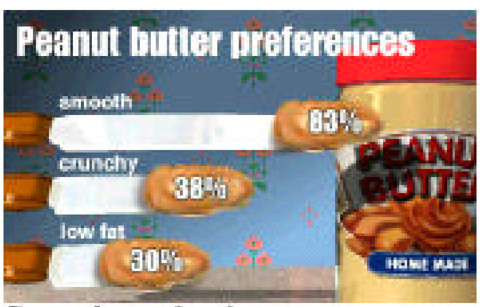
8.



Source: Bruskin/Goldring Research for Sears, Roebuck

Weekly yard care

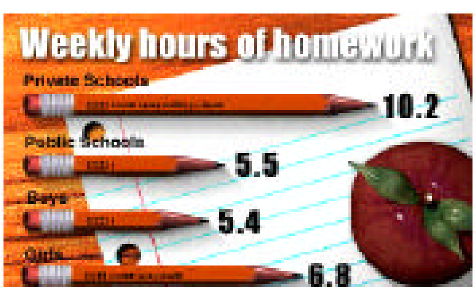
9.



Smooth or chunky

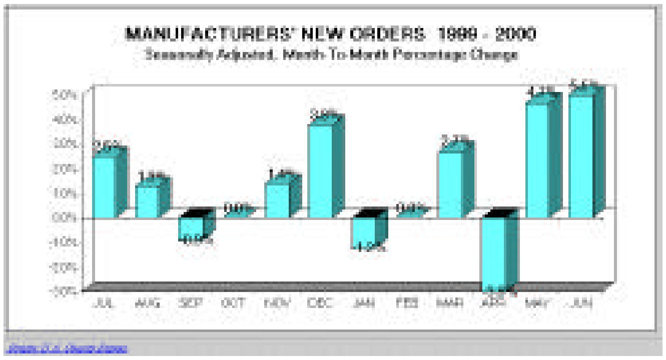
Nearly 60% of Americans of all ages eat peanut butter at least once a week. In the past three months, women ate the following kinds of peanut butter 82% smooth, 39% reduced fat, 36% chunky and 30% old fashioned or natural. Men liked smooth the most 74%, chunky 42%, old-fashioned/natural 40 and reduced fat 31%.

10.

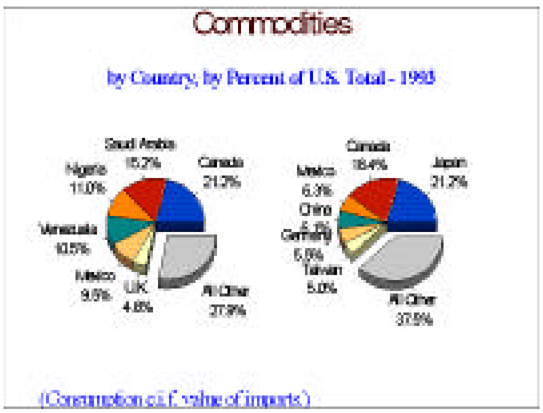


Source: The State of Our Nation's Youth 1998-1999, Horatio Alger Association

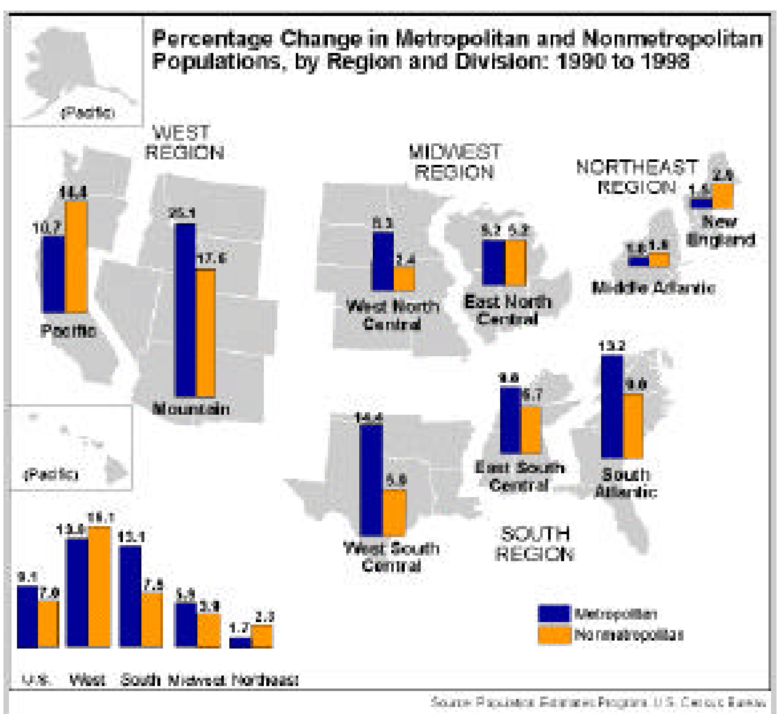
11.


MANUFACTURERS' NEW ORDERS 1999 - 2000
Seasonally Adjusted, Month-To-Month Percentage Change
JUL
AUG
SEP
OCT
NOV
DEC
JAN
FEB
MAR
APR
MAY
JUN

12.


Commodities
by Country, by Percent of U.S. Total - 1993
Saudi Arabia 15.2%
Canada 21.7%
Nigeria 11.0%
Venezuela 10.5%
Mexico 9.5%
UK 4.8%
All Other 27.9%
Canada 18.4%
Japan 21.2%
Mexico 6.3%
Taiwan 5.0%
All Other 37.5%
(Consumption c.i.f. value of imports)

13.


Percentage Change in Metropolitan and Nonmetropolitan Populations, by Region and Division: 1990 to 1998
(Pacific)
WEST REGION
MIDWEST REGION
NORTHEAST REGION
SOUTH REGION
10.7
14.4
Pacific
25.1
17.6
Mountain
West North Central
5.2
5.2
East North Central
New England
Middle Atlantic
13.2
South Atlantic
East South Central
14.4
West South Central
U.S.
West
South
Midwest
Northeast
Metropolitan
Nonmetropolitan

Project on Deceptive Displays of Data

The purpose of this project is to help you become aware of misleading graphic displays. For this assignment you are to collect an example of a misleading graph from a magazine or newspaper. You are to identify the source of your graph, and explain the reasons why you believe that it is deceptive. Finally, include your ideas on how the graph should be redone so that it better represents the data.

Your project should be done carefully and neatly and should have the following parts.

- A cover sheet that includes the title of the project, your name, and the date.
- The graph you are evaluating citing your source.
- Several paragraphs that address the ways that the graph may be misleading and your suggestions for how the graph should be redone so that it better represents the data. Use the questions below to help guide your thinking.

A Checklist for Visual Displays of Data

1. Is the message of interest clearly evident?
2. Does the graph have a title?
3. Is the purpose of the graph clearly evident?
4. Is the source of the data identified either with the graph or in an accompanying article?
5. Is the information from a reliable, believable source?
6. Is everything clearly labeled?
7. Do the axes start at zero?
8. Do the axes maintain a constant scale?
9. Are there any break in the numbers on the axes that may be easy to miss?
10. For data involving money, have the values been adjusted for inflation?
11. Is there extraneous information cluttering the picture or misleading the eye?
12. Does the number of dimensions depicted in the display exceed the number of dimensions in the data?

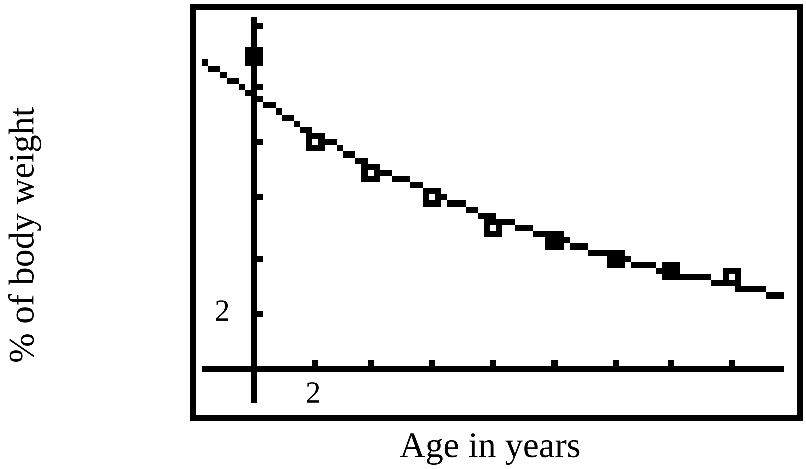
Tutfte’s Principles of Graphical Integrity

- Representation of numbers on a graph should be directly proportional to the numerical quantities represented.
- Graphs should be very clearly labeled and explanations written out on the graphic itself.
- Graphs should emphasize the data variation and not the graphical design.
- Graphical displays involving money over time should use deflated and standardized units. The data should be adjusted for inflation.
- The number of variable dimensions shown in the graph should not be greater than the number of dimensions in the data.
- Graphs must not give data out of context.

Predicting Outside the Domain

The graph below shows the relationship between age and brain weight as a percent of body weight. The points are actual values, and the curve is the best fit function for those values.

Brain weight as a % of body weight



1. As age increases what happens to the brain weight percentage?

2. Estimate the brain weight as a percent of body weight for a 9 year old child using the best fit curve.

3. A newspaper prints a story about the data and best fit curve relating age and brain weight. The headline reads, “*New brain research finding: The brain disappears as a person ages. A 75 year old is almost brainless*!”

   a. Is this a reasonable statement?

   b. Does the best fit curve support this statement?

   c. Is the curve a good model for this relationship? Explain.

Core Learning Goal 3:
Data Analysis and Probability

III. Probability and Simulations

Indicator 3.1.3

Using data, the students will determine the experimental or theoretical probability of an event.

Notes for Teachers

For many situations, specific outcomes cannot be predicted with certainty. Will the plane arrive on time? Will my luggage be lost? Will you be late to school tomorrow? Over the long run, however, we can establish how frequently these things happen. This relative frequency over the long run is called the probability of an event—it is the measure of chance or likelihood that the event will occur again. Students need to understand this concept of relative frequency. Students also need to know how to estimate probabilities from real data. The activities *Understanding Probability*, *Tumbling Thumb Tacks*, *Using Results of a Study*, and *Collecting Your Own Data* help students develop these concepts.

What is the difference between experimental probability and theoretical probability? How are these ideas related? Students need to understand this difference. They also need to experience the law of large numbers; that is, the more times an experiment is conducted, the closer the experimental results come to the theoretical results. The activity, *Spinning: From the Experimental to the Theoretical*, will help students with these ideas.

A very important part of this particular section is the concept and application of simulations. Simulation is a modeling technique used to help people answer questions about real situations by using probability experiments that are similar to the real situations. These are the basic steps in the simulation process:

- State the problem.
- State the assumptions.
- Select a model to generate outcomes.
- Simulate many repetitions.
- Analyze the results, and state the conclusion.

Students should conduct many investigations using simulations. Coins, spinners, and number cubes should be used first, followed by calculator or computer random number generators. The calculator and computer have the advantage of speed and ease, but the hands-on experiences are fundamentally important in helping students see that estimating probabilities is a dynamic process with variability in the outcomes.

The first three simulation activities—*Figuring Free-Throws*, *The Phone Tree with Built-In Back-Ups*, and *Thinking About Three*—involve a probability of 0.5. These can easily be modeled using coins or number cubes. The activity, *A Test Simulation*, involves a probability of 0.2 and would work well with a spinner or a calculator. *Figuring Free-Throws Again* involves a probability of 0.67 and either number cubes or a calculator would be easiest.

Another important concept in this section is that of expected values. Students need to know how to find and use expected values. A simple example of such a problem involves determining the number of boxes of cereal you would need to buy on average in order to get all three prizes that a cereal company is offering. Assume that each box of cereal has one prize and the prizes are distributed evenly among the boxes. Simulate this situation by generating numbers from 1 through 3, and count the number of digits you have to generate before you have a 1, a 2, and a 3. Doing this 30 times gave the following results: 7, 3, 5, 5, 5, 7, 4, 7, 8, 4, 4, 3, 4, 4, 5, 4, 7, 6, 4, 14, 3, 3, 5, 3, 6, 4, 7, 4, 5, 3. This list of data shows that for the first trial 7 digits were randomly generated before a 1, a 2, and a 3 appeared. For the next trial, 3 digits were generated showing a 1, a 2, and a 3. The list continues showing the results of 30 trials. The mean of this data is 5.1, or on average we can expect to have to buy approximately 5 boxes of cereal to get all three prizes.

The theoretical average number of boxes purchased is 5.5. In general, the theoretical waiting time to get *n* equally likely prizes is

$$\frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \ldots + \frac{n}{3} + \frac{n}{2} + \frac{n}{1}.$$

This formula is based on the fact that if the probability of an event is *p*, the average or expected number of trials before it occurs is $\frac{1}{p}$. So if we have 1 of the 3 prizes already, then the probability of getting a new prize in the box of cereal is $\frac{2}{3}$. This means that, on average, we would have to purchase $\frac{3}{2}$ additional boxes to get a new prize. Problems dealing with this concept include *Simulation with an Unknown Number of Trials* and *Calling for the Newspaper*.

Understanding Probability

Consider the following events and estimate how likely each is to occur. Use the following scale to help you. A "0" means that the event will never happen, and a "1" means that the event is certain to happen.

Impossible____________ | ___________ | ___________ | ____________Certain
0 .25 .5 .75 1

1. You will get an A on your next math quiz.
2. You will have chicken for dinner tonight.
3. The sun will set this evening.
4. It will rain sometime this week.
5. It will snow in Maryland this July.
6. A dog can live without water for one month.
7. The next baby born at Union Memorial Hospital will be a girl.
8. You will have scrambled eggs for lunch one day this week.
9. You will be absent from school at least one day during this school year.
10. A Republican will win the next presidential election.

Probability and Expected Patterns
Directions for Teachers

In this activity students will see that over the long run we can expect certain outcomes.

In this case where students essentially guess the answers to a 10-item true/false test, we can expect few perfect papers but many papers with scores of 4, 5, or 6. This is a fun activity that will generate high interest and involvement. It is an excellent introduction to the unit on probability and simulations for students who do not need something as basic as *Understanding Probability*.

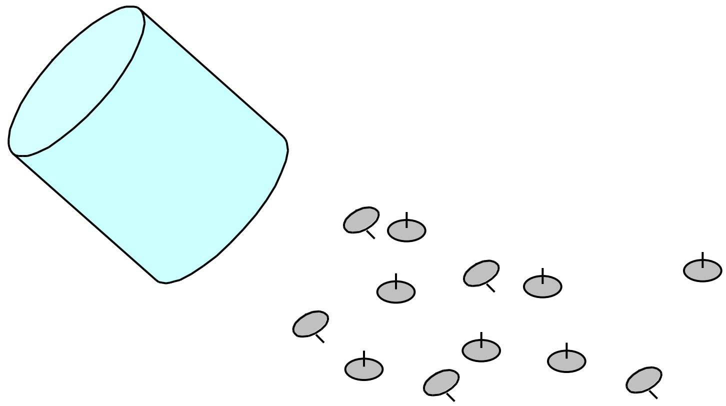
1. Have the students number their papers 1 through 10. Each student should use a coin to determine whether the answer to each item is true or false. Heads is true and tails is false. Have a supply of pennies available in case any student needs a coin.

2. Now that the students have the answers, the fun part begins. You are to read the questions and have the students grade their own papers. Be as creative as you want to be. A set of questions is given below, but you are encouraged to personalize this for your students. Note that as you read the questions, you also give the correct answer.

    1. The Washington Monument is located in New York City. *false*
    2. Mrs. Yourname is the best teacher in the school. *true*
    3. Mr. Yourname is the better looking than Brad Pitt. *true*
    4. Ice cream is one of the basic four food groups. *false*
    5. Everyone in this class loves mathematics. *true*
    6. The state of Maryland has declared that winter break will be one month long. *true*
    7. The school cafeteria is serving steak and lobster for lunch tomorrow. *true*
    8. No one in this room watches MTV. *false*
    9. You are getting a BMW for your birthday. *false*
    10. Mrs. Yourteacher never gives homework. *true*

3. Have the students grade their papers to determine the number they got correct. Create a line plot on the blackboard so that the students see the distribution pattern for the class. You should expect to see a fairly normal distribution pattern with few scores of 10 or 0, and more scores or 4, 5, or 6. The important point of doing this exercise is to have students understand that certain events are going to have particular expected values over the long run. To emphasize this point have the students flip their coins 10 more times and count the number of heads. Their values should then be plotted on a new line plot. This will allow the students to see that basically the same patterns are repeated.

Tumbling Thumb Tacks



Shake 25 identical thumb tacks in a cup and gently dump them onto a flat surface. Count the number of tacks landing point-up and the number landing point-down. Record that information in your data table.

Number of Tacks Pointing Up	
Number of Tacks Pointing Down	
Total	

Consider the data you collected as you answer these questions.

1. Which is more likely—point-up or point-down?

2. Estimate the probability of tossing one tack and having it land point-up.

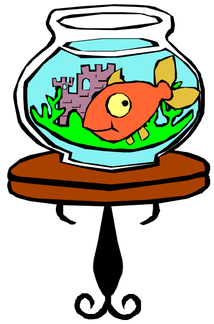
3. Combine your data with three other students so that you have data on a total of 100 tacks. Record the totals in the following data chart.

Number of Tacks Pointing Up	
Number of Tacks Pointing Down	
Total	

4. Use the combined data to determine the probability that a tack will land point-up.

5. Compare your results to the results of other groups in your class.

Using the Results of a Study



Suppose that 40 students from your school were interviewed. Of the randomly selected students, 18 were boys and 22 were girls. Fourteen of the boys and 15 of the girls have family pets. The information is displayed in the table:

	Girls	Boys	Total
Has a family pet	15	14	29
Does not have a family pet	7	4	11
Total	22	18	40

Another student is randomly selected from the student body.

1. Estimate the probability that the student is a boy.
2. Estimate the probability that the student does not have a family pet.
3. Estimate the probability that the student has a cat.
4. If the randomly selected student is a girl, estimate the probability that she has a family pet.
5. Suppose there are 1,600 students in the school. How many would you expect to have a family pet?

Collecting Your Own Data

Interview a number of students in your school to see how many have ever been to an Orioles game and how many have been to a Ravens game. Ideally these students should be randomly selected, but you may have to settle for a convenience sample. Record the numbers in the following table.

	Has been to a Ravens game	Has not been to a Ravens game	Total
Has been to an Orioles game			
Has not been to an Orioles game			
Total			

Consider another student in your school whom you did not interview. Using the data you collected from your interviews complete the following items.

1. Estimate the probability that the student has been to an Orioles game.

2. Estimate the probability that the student has been to a Ravens game.

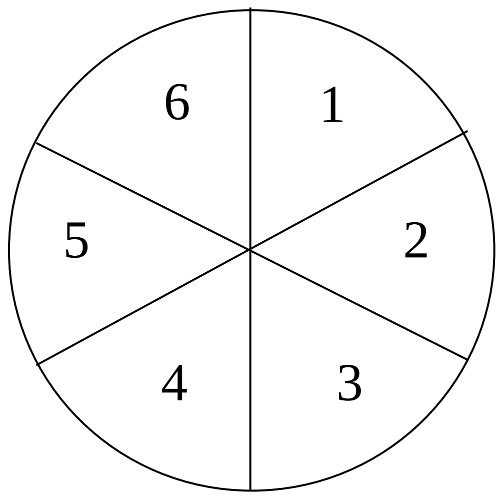
3. Estimate the probability that the student has been to both an Orioles and a Ravens game.

4. Consider a group of 100 students from your school whom you did not interview. How many students would you predict have not been to an Orioles game or a Ravens game?

Spinning: From the Experimental to the Theoretical

In this activity you will explore the relationship between theoretical and estimated probability. Remember that, for an experiment with equally likely outcomes, the probability of an event E is defined to be

$$P(E) = \frac{\text{number of favorable outcomes}}{\text{total number of outcomes}} = \frac{f}{n}.$$



Consider the "spinner" above. Since the six numbered sections are about the same size, then the six possible outcomes should be equally likely. Use the definition of theoretical probability to complete the following items.

1. Find the P(*spinner landing on 4*).

2. Find the P(*spinner landing on an odd number*).

3. Find the P(*spinner landing on 8*).

Spin the spinner 30 times and record your results in the table below.

	1	**2**	**3**	**4**	**5**	**6**	**Total**
Number of Times Observed							30

4. Find the fraction of 4's observed in the total sample. Compare this to your answer for the question number 1.

5. Find the fraction of odd numbers in the total sample. Compare this to your answer for question number 2.

6. Your answers for items 4 and 5 are estimates of probabilities. How would those estimates compare to the theoretical probabilities in questions 1 and 2 if we had data on 10,000 spins rather than 30 spins?

ESP—Do You Have It?

Have you ever wondered whether you have extrasensory perception? Scientists at Duke University have been studying this topic for many years. Is there really any such thing as ESP? Carry out the following experiment and see what you think.

Make a set of 40 cards using the following four symbols. Make 10 cards for each symbol.

Work with a partner to do this experiment. Have your partner face away from you. After shuffling the cards well, turn them over one at a time and concentrate on the symbol on the card. Ask your partner to identify the symbol on the card you are observing. Record whether the answer is right or wrong. Be sure to *not* indicate to your partner whether the response was correct or not. Continue the experiment until you have gone through the entire set of cards.

	Tally	Total
Right Answer		
Wrong Answer		
Total		

1. What is the probability of getting one response correct simply by guessing?

2. With 40 cards, how many correct responses would you expect to have just by guessing?

3. Do you think your partner has ESP? Use mathematics to justify your answer.

4. Suppose you were to repeat this experiment with 100 trials. Predict how many responses would be correct?

Simulation

Simulation is a modeling technique that is used to help people answer questions about real situations by using probability experiments that are similar to the real situation. It is important understand the basic steps involved in conducting a simulation. A simple example is given to illustrate the process.

Problem 1: What are the chances that when you toss a coin 7 times you will have a run of at least 3 consecutive heads or 3 consecutive tails?

Step 1 State the Problem

If you toss a coin 7 times, what is the probability that you will have a run of at least 3 consecutive heads or 3 consecutive tails? (Note: It is important to have all of the information available and to understand the objective of the simulation.)

Step 2 State the assumptions

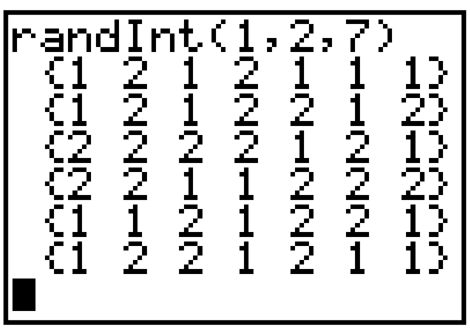
When you toss a coin, a head or a tail is equally likely to appear. Tosses are independent. In other words, the outcome of any toss will not influence the next toss.

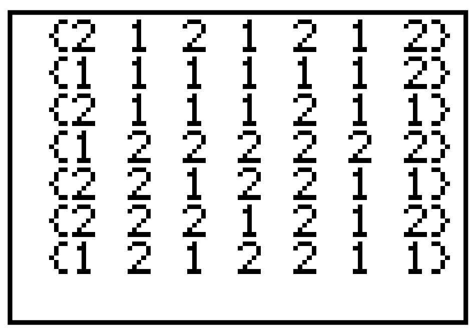
Step 3 Select a Model to Generate Outcomes

Let the digit 1 represent heads and the digit 2 represent tails. Use a random number generator or a random number table for getting digits.

Step 4 Simulate Many Repetitions

For this problem use the random integer feature of the TI-83. (Note: randInt is found by pressing MATH then PRB and selecting 5:randInt. The numbers 1, 2, 7 indicate that you are generating 7 random integers ranging from 1 to 2.)





Each time you press the enter key you generate another repetition of the experiment. When you are estimating the probability of getting a run of 3 heads or 3 tails when a coin is tossed 7 times, the more repetitions you have, the better your estimate will be. The results displayed on the calculator screen show that out of 13 trials there were 7 times when there were 3 heads or 3 tails in a row.

Problem 2: What are the chances that when you toss a coin 7 times you will get exactly 3 heads? Use the random binomial feature of the TI-83 to generate 5 trials of 7 tosses at a time. Then each time you press [ENTER] another 5 trials are simulated. (Note: randBin is found by pressing [MATH] then PRB and selecting 7:randBin. The numbers 7, .5, 5 indicate that you are completing 7 tosses with a probability of success of 0.5 and you are generating 5 trials at one time.)

```
randBin(7,.5,5)
     {3 0 5 3 3}
     {2 5 4 3 4}
     {5 5 2 5 5}
     {3 4 5 4 4}
     {3 4 4 6 4}
     {2 3 4 1 4}
```

Out of the 30 simulations generated by the calculator, exactly 3 heads appeared 7 times. So the experimental probability is 0.2333.

Problem 3: If you roll a die 5 times, what is the probability that you will get exactly one 6?

Method 1:

```
randInt(1,6,5)
     {4 3 5 3 6}
     {3 3 1 4 5}
     {4 4 3 2 3}
     {2 2 2 2 2}
     {5 3 4 4 2}
     {6 1 3 3 3}
```

```
     {1 4 4 1 2}
     {4 2 6 4 5}
     {3 4 4 3 6}
     {1 4 1 6 2}
     {1 6 5 2 6}
     {3 4 1 2 4}
     {5 3 6 4 6}
```

Since you rolled exactly one 6 in 5 out of 13 trials, the experimental probability is 0.3846.

Method 2:

```
randBin(5,(1/6),
5)
     {2 0 1 1 1}
     {1 1 0 1 2}
     {0 0 2 0 1}
     {2 2 0 1 0}
     {0 0 0 0 1}
```

Since you rolled exactly one 6 in 9 out of 25 trials, the experimental probability is 0.36.

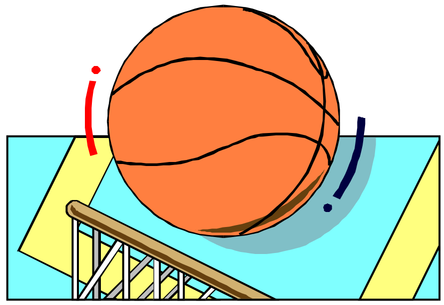
Try a few more problems on your own.

4. If you roll a die 7 times, what is the probability that you will get all even numbers?

5. If you roll a die 3 times, what is the probability that you will get all odd numbers?

6. If you toss a coin 10 times what is the probability that you will get 3 or fewer heads?

Figuring Free-Throws



Freddy completes 50 percent of his free-throws. He generally shoots 9 free-throws in a game. Use a simulation model to estimate the following. Run at least 50 trials in the simulation.

1. Estimate the probability that Freddy will complete all of his free-throw shots in any one game.

2. Estimate the probability the he completes exactly 4 of his free-throw shots.

3. Estimate the probability that he completes at least 5 of his free-throw shots.

4. In any one game, what is the most likely number of free-throw shots that Freddy will make?

5. Do you think that the Terps should recruit Freddy?

Thinking About Three

Tina and Tom have just gotten married, and although they do not want to start a family just yet, they are planning on having three children eventually. They have already started saving money for college for their future children, but they need some help in thinking about saving for future weddings of possible daughters. Use a simulation to help them estimate the following probabilities. Complete at least 60 trials of families with three children.

1. Estimate the probability that Tina and Tom will have three girls.

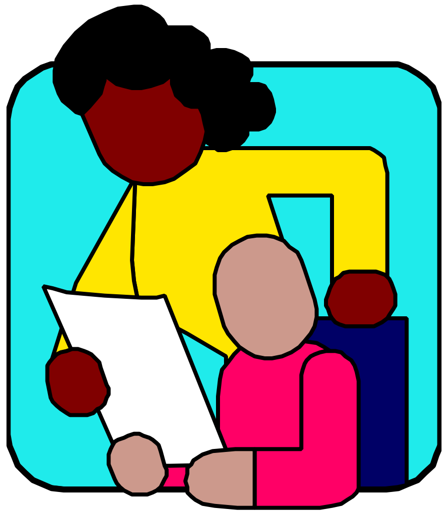
2. Estimate the probability that they will have two girls and one boy.

3. Estimate the probability that they will have one girl and two boys.

4. Estimate the probability that they will have three boys.

5. What is the sum of your answers for items 1 through 4? Do you think your answer makes sense? Use mathematics to explain why or why not.

A Test Simulation



Thomas was taking a multiple-choice test with 10 questions. Each question had 5 possible answers, *a* through *e*. Use a simulation to answer the following questions.

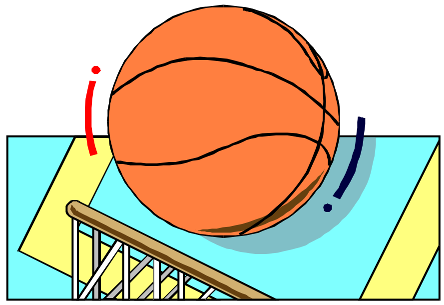
1. What is the probability that Thomas would answer at least four of the questions correctly, if he guessed all the answers?

2. What is the probability that Thomas would answer at least seven of the questions correctly, if he guessed all the answers?

3. On average, how many questions would Thomas answer correctly simply by guessing?

4. How would your answers change if the test had only 4 possible answers instead of 5?

Figuring Free-Throws Again



You may remember Freddy from an earlier activity. Well, Freddy has brought his free-throw average up from 50% to 67%. Use a simulation model to answer the following questions. Complete at least 30 trials.

1. Explain how you could use the following items to simulate Freddy's free-throws: a spinner, a die, a random number table, and a calculator or computer.

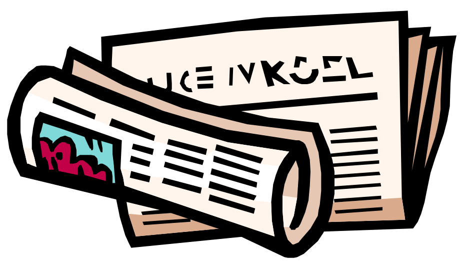
2. The Terps and the Blue Devils are tied up with no time left on the clock. Freddy was fouled and gets two shots. What is the probability the he will make at least one shot to win the game for the Terps?

Simulation with an Unknown Number of Trials

A certain breakfast cereal includes one of seven possible prizes in each box. There are an equal number of each prize so that the probability of getting any one of the seven prizes is the same. On average, how many boxes of cereal would you need to buy to get all seven prizes.

This problem is different from the simulation problems that have been done so far. For this problem you need to simulate buying cereal until you have all seven prizes. One way to do that is to let the digits 1 through 7 represent the seven different prizes. Then generate integers from 1 to 7 until all 7 numbers have appeared. This would be considered one trial. If you had generated 17 numbers before all 7 digits appeared, your first trial indicates that you would have to buy 17 boxes. You would repeat this a large number of times. The average of these numbers is your estimate of the expected number of boxes of cereal you will have to buy in order to get a complete set of prizes.

Calling for the Newspaper



Do you get calls around dinner time asking you whether you want to have the local newspaper delivered? Suppose the probability that an exploratory call like that results in a subscription to the paper is 15%. Conduct a simulation to find answers to the following questions. Assume that the outcome for any one exploratory call is independent of outcomes for other calls that have been made.

1. Estimate the average number of calls made before getting an order.

2. For your simulation, what was the greatest number of calls made, including the first successful call?

3. Telephone solicitors are paid $7.00 per hour and they make on average 30 calls per hour. If they receive a bonus of $2.00 for every order, what would they earn per hour on average with the bonus?

Core Learning Goal 3:
Data Analysis and Probability

IV. Experimental Design

Indicator 3.1.1

The students will describe how they would do an investigation, select an investigation, and defend their choice. Students will consider simple random sampling (SRS) techniques that may include sampling size, bias representation, and randomness.

Notes for Teachers

Students need to understand the concepts of a simple random sample, sampling size, and biased representation in order to be prepared for the High School Assessment. A sample is random if both these conditions are true: each member of the population is equally likely to be chosen, and the members of the sample are chosen independently of one another. It is often not possible to pick a simple random sample, however, and in such cases bias may be introduced. Bias may be present in results because those results are based on a voluntary response sample. Bias might also be due to undercoverage, nonresponse, or poorly worded questions. In addition to these ideas, students need to understand that in order to be confident about the results of a survey or an experiment, they must conduct enough trials or they must survey enough randomly selected people to get good results. The larger the number of trials or people surveyed, the more confident students can be in the results. In general, these concepts are not treated in a formal way in these activities, but students need to have a fundamental understanding of these ideas. Activities that emphasize these concepts are *Random Samples*, *Simple Random Samples*, *Random Digit Tables*, and *Random Rectangles*.

Random Samples

A mini theater has 100 seats as shown in the given diagram.

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
J1	J2	J3	J4	J5	J6	J7	J8	J9	J10

All seats were filled during the 2:00 showing of a movie one Saturday. The theater manager wants to estimate the proportion of females in this audience. Instead of counting the number in the entire audience he will use a **sample** of 10 people from the audience.

1. How could he estimate the proportion of females in the audience from a sample of 10 people?

2. The manager is considering the following techniques for choosing the sample:
   - Randomly select a row, the sample will consist of the 10 people in that row.
   - Randomly select a seat number from 1 to 10, the sample will consist of the 10 people in that seat in each row.

   Would you expect the proportion of females in each of these samples to be a good estimate of the actual proportion of females? Explain.

   Examine the diagram that shows the location of the females in the audience.

3. List the samples that result using each of the methods described in # 2.

4. What is the proportion of females in each of these samples?

5. What is the proportion of females in the audience?

6. Is the sample proportion of females for each technique a good estimate of the actual proportion of females? Explain.

7. The manager has the audience write their gender on the back of their ticket. He has decided to choose the sample by randomly selecting 10 tickets from a bag containing all of the tickets

   a. Will the sample proportion of females using this technique be a better estimate of the actual proportion than those in number 2? Explain.

   b. Obtain several samples using this technique. Calculate the sample proportion of females for each sample. Compare the sample proportions to the actual proportion of females.

The diagram below shows the location of the females in the audience.

	F	F			F	F	F		
	F	F			F	F	F		
	F	F			F	F	F		
	F	F			F	F	F		

Simple Random Samples

A sample is a **simple random sample** if
- each member of the population has an equal chance of being selected and
- each possible sample is equally likely.

1. Decide whether or not the following sampling methods produce a simple random sample from a class of 30 students. Explain the reasoning behind your decision.
   a. A teacher wants to select five students from the class. She selects the first five students that enter the room.
   b. A teacher wants to select ten students from the class. She lists students in alphabetical order, then selects every third student.
   c. A teacher wants to select six students from the class. She writes each student's name on an index card, places the index cards in a box, mixes the cards, then chooses six cards from the box.

2. Occasionally, random sampling yields a sample that is not *representative* of the population. Suppose there are fifteen boys and fifteen girls in a math class. Each student's name is placed in a hat and the names are thoroughly mixed. Seven names are drawn and all names correspond to the boys in the class.
   a. Did the sampling method produce a simple random sample? Explain.
   b. Is this a *representative* sample? Explain.

3. For each of the following sampling methods, identify the groups in the population that are *underrepresented*.
   a. To obtain a sample of households, a consumer reporter dials numbers taken at random from a telephone directory.
   b. A car manufacturer wishes to survey a sample of drivers, so he randomly selects the names of car owners from a list of vehicle registrations.
   c. A college professor wants to know what percentage of young adults, ages 18 to 22, consider education a top priority. He obtains a list of all students on campus from the registrar and randomly selects names from the list.
   d. A radio station wishes to examine the proportion of its listeners who voted in the last presidential election. They conduct a poll by asking listeners to call the station.

4. The student government at Eastern High School wanted student ideas for social activities. The group sent out a questionnaire to a simple random sample of 600 students. The results are shown in the table below.

Idea	Dance	Ice Cream Social	Movie Night	Carnival
No. in Favor	120	80	150	250

   a. What proportion of students in the survey favored the Carnival?

b. What proportion of the entire student body would you expect to favor the carnival?

c. If Eastern High School has a population of 2400 students, how many would you expect to favor the Carnival?
d. A student from Eastern High School is randomly selected and asked which social activity he prefers. What is the probability that he favors the school dance based on this sample?
e. Suppose the 600 students in the sample were *not* selected randomly. Would it be fair to say that 25% of the student body favors Movie Night? Explain.

Random Digit Tables

Examine the following table of random digits.

62207	96845	33122	61147	93253	60200	85048	61922	37863	20812
91294	71538	27054	33696	24444	70998	51609	69031	36872	55220
12148	31711	35563	05855	53337	21329	52349	77117	73675	24707
79180	65454	79577	24794	73456	45187	81324	56924	58063	79666
24439	91645	83361	59133	13884	83711	95824	86031	69907	38546
04835	99007	56062	87960	01512	90025	93472	74303	05549	72506

1. There are 300 digits between 0 and 9 (inclusive) in the table.
    a. About how many 5s would you expect to find? Check your answer by counting the number of 5s that appear.
    b. About what percentage of digits in the table would you expect to be a 5?
    c. About what percentage of digits in any large table of random digits from 0 to 9 will be even?
    d. About what percentage of digits in any large table of random digits will be less than 5?
    e. About what percentage of digits in any large table of random digits will be either a 2 or a 5?
    f. About what percentage of the 1s in a large table of random digits will be followed by a 2?

2. a. How could you use a table of random digits to select 2 people from a group of 10 people?
    b. Select 2 people from a group of 10 people using the random digit table.

3. a. How could you use a table of random digits to select 5 people from a group of 100 people?
    b. Select 5 people from a group of 100 people using the random digit table.

Table of Random Digits

Line									
1	21265	26810	00264	93115	59105	11038	92439	94202	00797
2	06960	94142	53813	81802	74430	09838	20729	40293	59497
3	31504	13614	86475	98885	90257	83328	98839	56027	31745
4	72246	60999	65442	60622	82496	78673	35934	49230	27814
5	99551	89597	34571	48599	94791	15772	77290	17687	59867
6	88710	37677	73661	33286	35149	20170	89056	21718	21260
7	06242	49406	71026	70376	97795	66224	72519	51421	27991
8	32279	92225	33234	56097	56680	41699	14595	43766	92961
9	24106	67221	31853	50663	73972	87862	02930	88036	45091
10	16417	34657	46273	48811	64709	37474	36522	34383	08516
11	86971	75557	73474	51637	06986	90224	63563	10078	99682
12	08214	68047	50838	12498	10263	94204	42062	01963	46070
13	78843	99626	63150	54996	48418	45656	45816	08191	75077
14	04433	04927	72273	50982	58022	70262	63184	98328	80167
15	05916	03851	27818	53023	69185	53434	20701	98342	98926
16	03447	89559	38909	56375	64313	36477	05723	46874	22370
17	34040	37836	49754	75009	73087	28956	14573	38701	37364
18	43303	36733	51106	83337	69375	71377	99451	00583	11982
19	74081	66356	39164	65919	50580	92842	89446	01979	44352
20	24834	99106	81335	23243	18630	98839	87682	54381	50074
21	32155	45065	10367	94236	13383	25581	95417	47937	90486

Random Rectangles

You will be given a worksheet containing 100 numbered rectangles.

1. Briefly examine the areas of the rectangles on the worksheet provided.

   a. Guess the average or mean area of these rectangles.

   b. Draw a line plot showing the class guesses.

2. a. Choose five rectangles that you think are representative of these rectangles and write their areas in the space provided.

rectangle number					
rectangle area					

   b. What is the average or mean area of the rectangles that you chose?

   c. Draw a line plot showing the class averages.

3. a. Choose five rectangles using a random number table or generator and write their areas in the space provided.

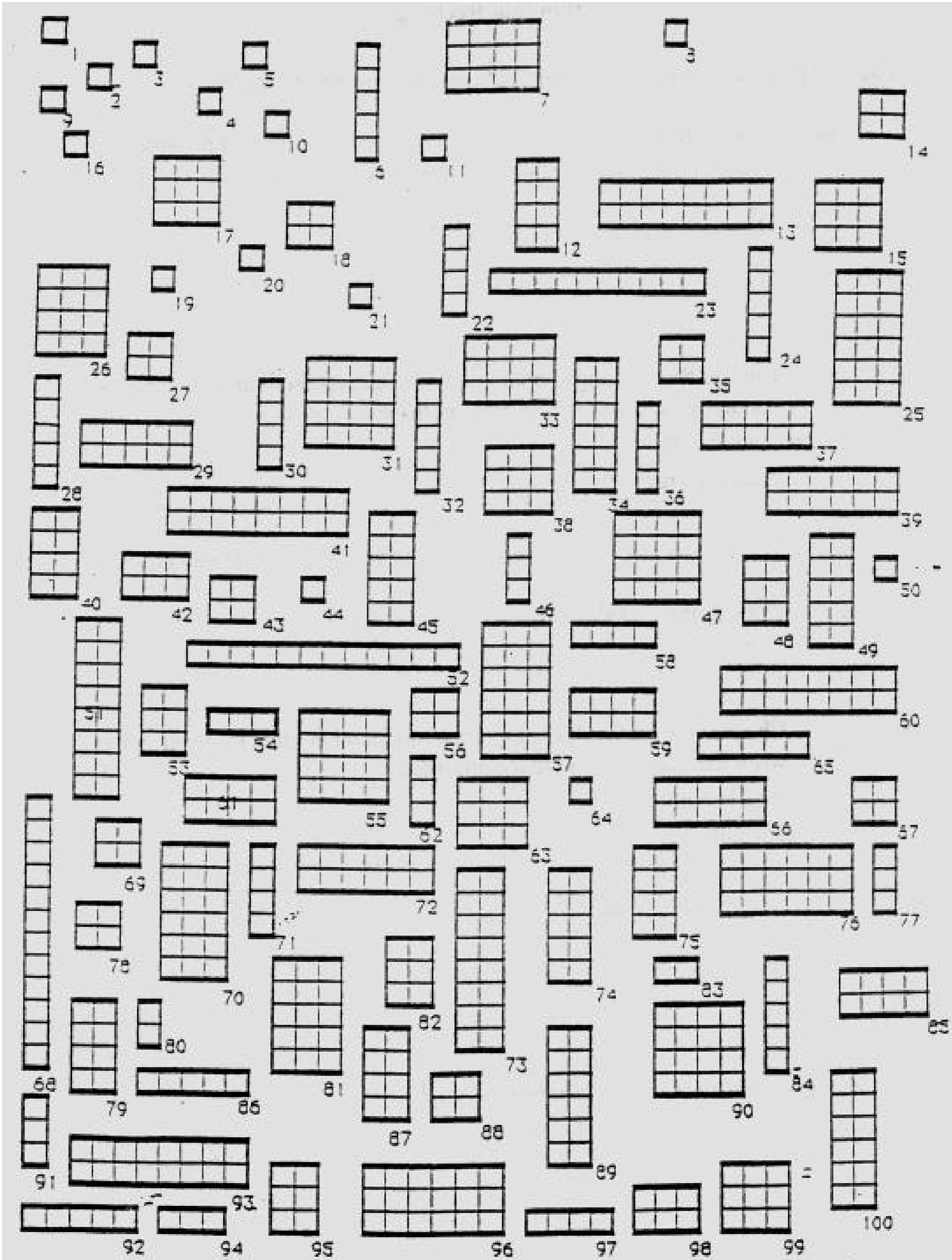
rectangle number					
rectangle area					

   b. What is the average or mean area of the randomly chosen rectangles?

   c. Draw a line plot showing the class averages.

4. Compare the class means using each sampling technique above.

5. Compare sampling mean areas to the actual mean area of the rectangles.



Core Learning Goal 3:
Data Analysis and Probability

V. Line of Best Fit

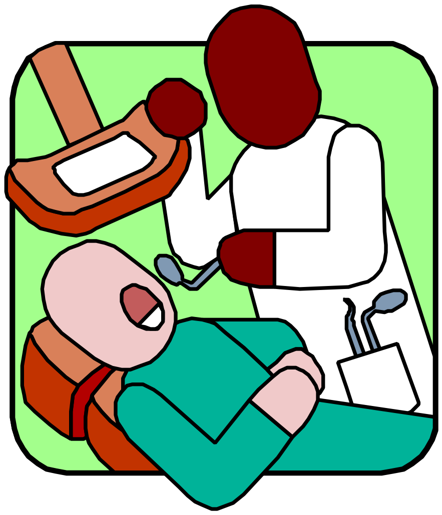
Indicator 3.2.2

The students will demonstrate their understanding of the process by finding a line of best fit and by using it to make predictions and/or interpret data (slope and intercepts) or by using a curve of best fit to make a prediction.

Notes for Teachers

Students need to be able to determine a line of best fit and use that line to answer questions or make predictions. They need to be able to interpret the y-intercept and the slope in terms of the context of the problem. They need to totally understand the process and the relationships among the table of values, the symbolic rule, and the graph. In addition to determining the line of best fit, students need to be able to make sense out of a curve of best fit. They should be able to interpret data and make predictions using the table, the equation, or the graph.

Dentists for the Future



The Number of Active Dentists in the United States 1980 – 1994
(In thousands)

Year	**1980**	**1985**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**
Number	**121**	**136**	**144**	**147**	**149**	**152**	**154**	**157**

Everyone values good health—including good dental health. Good dental health depends in part on the availability of good dental care. Will there be enough dentists in the future to keep our teeth healthy?

1. What sort of pattern do you notice in the table?

2. Make a scatter plot of the (*year*, *number*) data. Be sure to use an appropriate scale.

3. Draw a line of best fit on your scatter plot and use this line to predict the number of active dentists in the United States in the year 2010.

4. Determine an equation for your line.

5. What is the slope of your line? Explain the meaning of the slope in terms of the problem.

6. Can you use your line or your equation to determine whether the United States will have enough dentists in the year 2010? What are some other important considerations that might go into determining whether we will have enough dentists.

How Long Will Patients Stay?

The Average Length of a Hospital Stay
1989 – 1994

Year	89	90	91	92	93	94
Average Stay	6.5	6.4	6.4	6.2	6.0	5.7

Source: U.S. National Center for Health Statistics

1. Use your calculator to create a scatter plot for the data in the chart.

2. Explain the type of pattern that you see.

3. Use your calculator to determine the equation for a line of best-fit.

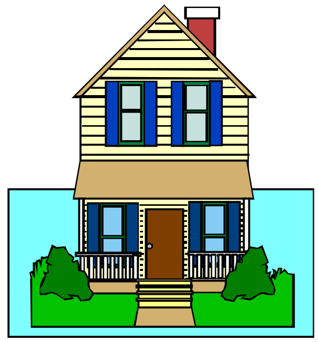
4. Graph this line on your scatter plot.

5. What is the slope of your line? Explain what the slope means in terms of the context of this problem.

6. Use this line to determine the average length of a hospital stay in the year 2000. Do you think that your answer seems reasonable? Call a local hospital and ask them what the average length of a hospital stay is.

7. Use your graph or your equation to predict the average length of a hospital stay in 2005. Explain how reliable you believe your answer is.

Homes for the Mentally Retarded



Over the last 25 years there have been major changes in the way our society cares for the mentally retarded. Today, mentally retarded men and women live and work in our communities. In the past it was more common for mentally retarded citizens to be cared for in residential facilities such as hospitals or institutions.

The Number of Mentally Retarded Citizens in Residential Facilities 1980 – 1993

Year	80	85	90	92	93
Number of Residents	148734	117101	94625	86498	73856

Source: Center for Residential Services and Community Living

This relationship can be modeled by the equation $y = -5395x + 147326$.

1. What type of function does this equation represent?

2. What is the slope? Give the meaning of the slope in terms of this problem.

3. What is the y-intercept? What does this mean in terms of the context of this problem?

4. This equation was determined using x to represent the number of years since 1980. So, for example, if x equals 3 then you will be calculating the number of residents in 1983. Complete the chart using the equation above.

Year	x-value	Number of Residents
1980		
	4	
	7	
1989	9	
1995		
2000		
	22	

5. Do you think this pattern will continue? Explain your reasoning.

Assessment Items

1. The following table shows the percentage of sugar in several cereals.

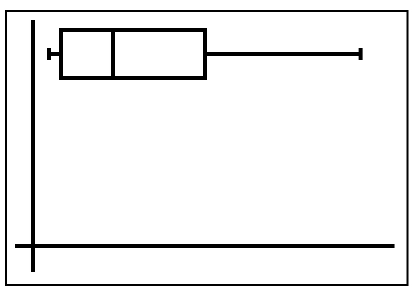
Brand of Cereal	Percent Sugar	Brand of Cereal	Percent Sugar
Sugar Smacks	56	Fruit Loops	48
Rice Krispies	7	Cocoa Puffs	33
Golden Grahams	30	Cracklin’ Bran	29
Raisin Bran	48	Honey Comb	37
Kix	4	Cocoa Crispies	43
Cookie Crisp	41		

Which measure of central tendency do you think best represents a “typical” value for this data? Explain.

2. A set of data contains a value that is significantly greater than the rest of the values in the set. Which of the following statements must be true?

   a. The mean of the data set is less than the median.
   b. The mean of the data set is the same as the median.
   c. The mean of the data set is greater than the median.
   d. The relationship between the mean and the median cannot be determined.

3. A boxplot for a set of data is shown on the display below.



Which of the following statements must be true?

   a. The mean of the data set is less than the median.
   b. The mean of the data set is the same as the median.
   c. The mean of the data set is greater than the median.
   d. The relationship between the mean and the median cannot be determined.

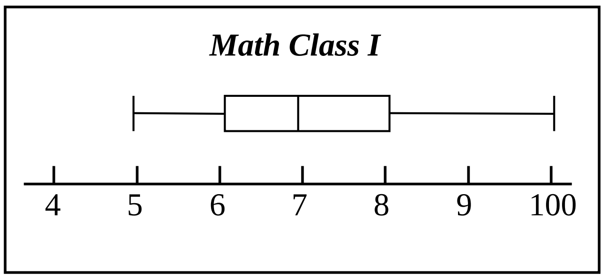
9. The table below shows the salaries of a small accounting business.

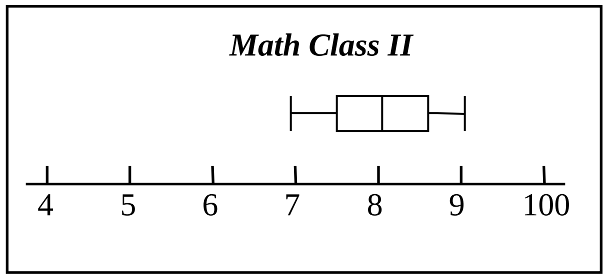
Title	Salary	# Employees
Clerk	$22,000	5
Jr. Accountant	$50,000	2
President	$270,000	1

a. If you were trying to recruit employees for this business which measure of central tendency would you advertise? Explain.

b. If you were interested in becoming an employee with this business which measure of central tendency would you want to know? Explain.

5. The following box plots show the distributions of test scores for two math classes. *Both* classes have 20 students.





Which class performed better on the test? Justify your choice.

6. Which of the following methods produce a simple random sample of students in your school?

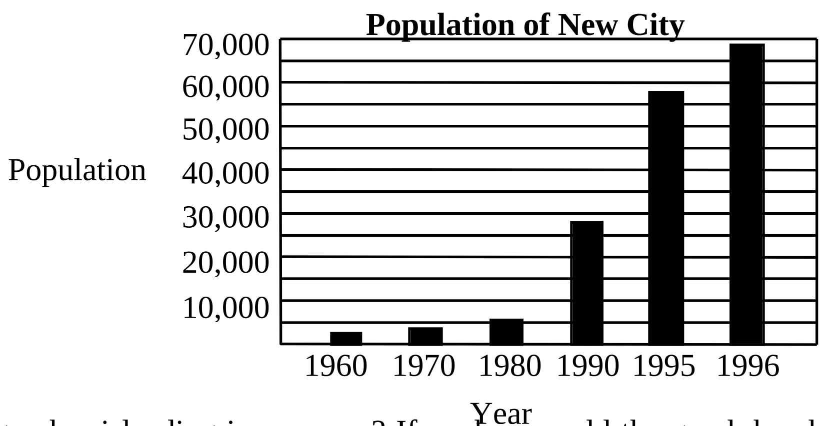
a. Place the names of all math teachers in a hat and draw three names. Select all students who are currently enrolled in the chosen math teachers' classes.
b. Give every student a raffle ticket with a unique number. Place part of the ticket in a box, then select 100 tickets. Announce the winning numbers, and choose students with those tickets.
c. Choose every fifteenth student entering the school building.
d. Obtain a list of student identification numbers, and choose all students whose number begins with the number 5.

7. A large bag of jelly beans contains four flavors: lemon, cherry, grape, and orange. The bag is thoroughly mixed and a handful of 50 jelly beans is drawn. The table below shows the flavors contained in the handful.

lemon	cherry	grape	orange
12	18	6	14

There are 400 jelly beans in the large bag. How many grape jelly beans would you expect to find based on this sample?

8. The graph below shows the growth in population of New City.



Is the graph misleading in any way? If so, how could the graph be changed to show the data more accurately?

9. The test scores of ten students in two classes are shown below.

Class A: 12, 29, 68, 80, 81, 87, 90, 90, 91, 99

Class B: 71, 72, 73, 74, 76, 80, 81, 84, 84, 90

Compare the performance of the students in Class A to the performance of those in Class B. Justify your conclusions using appropriate statistical measures.

10. A selected response test has four choices for each answer. If you realize that one of the answers to a question is not correct, what is the probability of guessing the correct answer?

a. 0 b. $\frac{1}{4}$ c. $\frac{1}{3}$ d. $\frac{3}{4}$

11. In a survey taken before an election, 3,000 people said they would vote **Yes** and 1,000 people said they would vote **No**. What is the probability a person will vote **No** based on this survey?

a. 25%

b. $33\frac{1}{3}\%$

c. $66\frac{2}{3}\%$

d. 75%

12. Of the five countries with the most computers in 1995, the United States had a significantly greater number of computers than the other four. Two magazines used

either the median or mean to report a "typical" number of computers in these countries.

- Magazine X stated that the "typical" number of computers was 30 million.
- Magazine Y stated that the "typical" number of computers was 14 million.

Which magazine used the mean as the "typical" number of computers in these countries? Explain your reasoning.

Use the following information to answer questions 13-15.

A baseball player's batting average is 0.400. Suppose she bats five times in a game. Use the numbers below from a random number table to simulate her times at bat.

72654 24625 78393 77172 41328 95633 55102 93408 10965 69744

Let the digits 0, 1, 2, and 3 stand for hits. Let the other six digits stand for no-hits.

13. How many trials are represented by the given numbers for this simulation?

    a. 1 b. 2 c. 5 d. 10

14. What number of hits is she *most* likely to get when she bats five times in a game based on this simulation?

    a. 0 b. 1 c. 2 d. 3

15. What is the probability that she gets at least one hit in five times at bat based on this simulation?

    a. 0.1 b. 0.2 c. 0.8 d. 0.9

16. Jane wants to obtain a *simple random sample* of 40 students from her school that has a total enrollment of 1600. She plans to have them fill out a survey form about course offerings. Below are some methods she is considering.

    a. Pick every 10th student that enters the school until 40 have been identified.

    b. Get a list of all of the students in the school from the guidance department and number the students from 1 to 1600. Use a calculator or computer to generate 40 different numbers.

    c. Get a list of all of the students in the school from the guidance department and number the students from 1 to 1600. Pick every 40th name on the list.

    d. Select 40 teachers and ask them to each randomly select one student to fill

    Identify which method is best.

- Describe why the best method is correct. Use mathematics to justify your answer.
- Describe why the other methods are not acceptable. Use mathematics to justify your answer.

17. Four teachers each surveyed a simple random sample of students from a class of 400 to determine how many would attend the school carnival. The results of their surveys are summarized in the chart below.

Survey Results

Name of Surveyor	Total Number of Students in the SRS	Number of Students Who Would Attend the Carnival
Mrs. Casper	60	43
Mrs. Flaherty	12	5
Mrs. East	25	15
Mr. Katra	10	8

- If you had to select the results of one survey to predict the number of students who would attend the carnival, which one would you use? Be sure to justify your answer using words or symbols or both.
- Predict the number of students will attend the carnival. Use mathematics to explain how you determined your answer. Use words, symbols, or both in your explanation.

18. Mr. Streagle used the field trip data in the table below to estimate the typical number of field trips per month for Carter High School.

Day	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Number of Field Trips	2	3	3	4	1	5	23	5	2	2

Which of these <u>best</u> represents the typical number of field trips per month?

a. 2 b. 3 c. 4 d. 5

Alternate format:

For the data given in the chart, give the mean, median, mode, and range. For each value, explain why or why not it would best represent the typical number of field trips per month. Use words or symbols to justify your answer.

Simulation Results

19. A basketball player makes 70% of her free throws in a long season. In a tournament game she shoots 5 free throws late in the game and misses 3 of them. Consider the results of a simulation based on her season record that are given in the chart.
    - What are all of the possible numbers that might appear in column #2?
    - Some fans thought that she missed three of the free throws because she was nervous. Do you agree with these fans? Use the results of the simulation along with words or symbols to justify your answer.

Trial Number	Number of Free Throws Missed Out of 5 Shots
1	3
2	2
3	1
4	1
5	2
6	3
7	1
8	1
9	3
10	1
11	3
12	2
13	1
14	0
15	0
16	3
17	0
18	1
19	1
20	3
21	2
22	2
23	2
24	2
25	2
26	0
27	0
28	0
29	3
30	1

20. Mathematics grades by gender for a particular class are given in the chart below.

Favorite Colors of Students in Two Kindergarten Classes

	Red	Orange	Yellow	Blue	Green	Purple
A.M. Kindergarten Class	6	4	3	4	2	1
P.M. Kindergarten Class	5	2	4	6	1	2

Which of these represents the probability that a P.M. kindergarten student likes yellow best?

a. 0.175 b. 0.200 c. 0.400 d. 0.571

21. The chart below gives average precipitation for Baltimore, Maryland for a number of years starting in 1974.

Year	**Precipitation (in inches)**	**Year**	**Precipitation (in inches)**
1974	41	1985	41
1975	62	1986	38
1976	44	1987	39
1977	40	1988	36
1978	45	1990	43
1979	61	1991	30
1980	32	1992	39
1983	58	1993	45
1984	38	1994	41

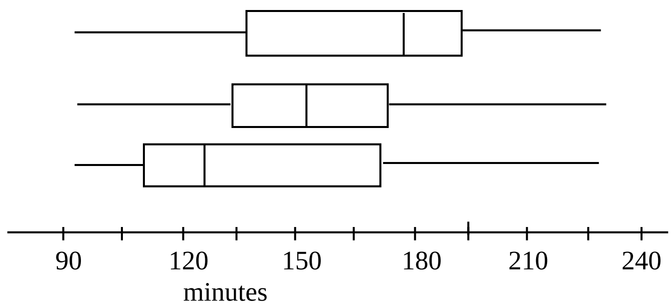
Source: **http://www.cdc.noaa.gov/cgibin/USclimate/rankall.calc.pl**

a. Create a boxplot using this data.

b. Describe the relationship between the mean and the median based on your boxplot. Use mathematics to justify your answer.

c. Explain what this relationship tells you about the average precipitation for Baltimore. Use mathematics and the original data from the chart to justify your answer.

22. The following parallel box plots show the effective duration of three over-the-counter pain relievers.



Which of the following statements is true?

a. The medians for the pain relievers are different but the means are the same.

b. All of the pain relievers are equally effective because they all have the same range.

c. Data for the middle pain reliever is approximately symmetric.

d. The basic shape of the data is the same for all three pain relievers.

Resources

Print Resources

Burrill, Gail, et. al. *Data Driven Mathematics: Exploring Regression*. Palo Alto, California: Dale Seymour Publications, 1999.

Burrill, Gail, et. al. *Data Driven Mathematics: Mathematics in a World of Data*. Palo Alto, California: Dale Seymour Publications, 1999.

Coxford, Arthur, et al. *Contemporary Mathematics in Context, A Unified Approach: Course 1, Parts A and B*. Chicago: Everyday Learning, 1997.

Coxford, Arthur, et al. *Contemporary Mathematics in Context, A Unified Approach Course 2, Part A*. Chicago: Everyday Learning, 1998.

Gnanadesikan, Mrudulla, et al. *The Art and Techniques of Simulation*. Palo Alto, California: Dale Seymour Publications, 1987.

Hopfensberger, P., et al. *Data Driven Mathematics: Probability through Data*. Palo Alto, California: Dale Seymour Publications, 1999.

Huff, Darrell. *How to Lie with Statistics*. New York: W.W. Norton, 1993.

Kranendonk, H., et al. *Data Driven Mathematics: Exploring Centers*. Palo Alto, California: Dale Seymour Publications, 1999.

Landwehr, James M, et al. *Exploring Surveys and Information from Samples*. Palo Alto, California: Dale Seymour Publications, 1987.

Landwehr, James M., and Anne Watkins. *Exploring Data*. Palo Alto, California: Dale Seymour Publications, 1986.

Newman, Claire M, et al. *Exploring Probability*. Palo Alto, California: Dale Seymour Publications, 1987.

Paulos, John Allen. *A Mathematician Reads the Newspaper*. New York: Basic Books, A Division of HarperCollins Publishers, Inc., 1995.

Scheaffer, Richard L. et al. *Activity-Based Statistics*. New York: Springer, 1996.

Tufte, Edward Rolf. *Envisioning Information*. Cheshire, Connecticut: Graphics Press, 1990.

Tufte, Edward Rolf. *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press, 1983.

Tufte, Edward Rolf. *Visual Explanations*. Cheshire, Connecticut: Graphics Press, 1997.

Internet Resources

Data Sites

http://www.stat.ncsu.edu/stated/data.html

Links to data sites such as the Sports Statistics on the Web, the Census Bureau, and DASL-a data and story library.

Chance Website

http://www.dartmouth.edu/~chance

Statistics newsletters, activities, real examples of misuse of statistics.

ExplorStat

http://www.stat.ufl.edu/users/dwack/

ExplorStat can be downloaded at this site along with Hypercard Player which is needed to operate the ExplorStat package. This package, developed through an NSF grant, has a variety of modules that are good for both student exploration and teacher demonstrations. The Dot Diagram module helps students build intuition regarding the relationship between data sets, corresponding numerical characteristics (mean, median, standard deviation, quartiles), and graphical displays including histograms and box plots.

Information Please Website

http://www.infoplease.com

Online dictionary, encyclopedia, almanac, and references containing current data sorted by category.

WWW Resources for Teaching Statistics

http://it.stlawu.edu/~rlock/tise98/onepage.html

Links to online resources for teaching statistics sorted by type.