

Resource Material Prototype

Mathematics GLE Resource Materials
Definitions and Examples for Grades K - 8

Data, Statistics, and Probability

January 2004

New Hampshire Department of Education
Rhode Island Department of Education
Vermont Department of Education

Resource Material Prototype

Overview

The purpose of these resource materials is to provide K – 8 educators with explanations and examples that facilitate understanding of the mathematics in the Grade-Level Expectations (GLEs)*. These resource materials are organized by content strand and by GLE “stems.” The definitions and explanations within each section are not alphabetized, but are organized to parallel the introduction of new concepts and skills within each GLE “stem” across grades. *The materials contained in this document focus on the Data, Statistics, and Probability strand and only address the NECAP GLEs that are common to all three states.*

Grade 3	Stem	Grade 4
M(DSP)–3–1 Interprets a given representation (line plots, tally charts, tables, or <u>bar graphs</u>) to answer questions related to the data, to analyze the data to formulate conclusions, or to <u>make predictions</u> .		M(DSP)–4–1 Interprets a given representation (line plots, tables, bar graphs, <u>pictographs</u> , or <u>circle graphs</u>) to answer questions related to the data, to analyze the data to formulate or <u>justify</u> conclusions, to make predictions, or to <u>solve problems</u> .
(IMPORTANT: Analyzes data consistent with concepts and skills in M(DSP)–3–2.)		(IMPORTANT: Analyzes data consistent with concepts and skills in M(DSP)–4–2.)

To facilitate access, each definition is coded (e.g., *DSP – 4*, the 4th definition for the Data, Statistics, and Probability strand). There are multiple tables of contents: 1) An overall table of contents on page 3 contains all the terms or phrases defined for the content strand in alphabetical order; and 2) A table of contents for each section in alphabetical order. In addition, if there is a word or phrase that you are unclear about within a definition, check the overall table of contents to determine if the word or phrase is defined in another location in this document.

The materials are divided into 2 sections: Section 1 includes definitions and examples for the GLEs related to Data and Statistics; Section 2 includes definitions and examples for the GLEs related to Counting Principles and Probability.

Section #	Grade Level Expectation “Stem”	NECAP GLE Stem Codes (Vermont codes) X represents grades K -8	Page Numbers
Section 1	Interprets a given representation... Analyzes patterns, trends, and distributions in a variety of contexts... Identifies or describes representations or elements of representations...	NECAP: M(DSP) – X – 1 (Vermont GE MX: 23) NECAP: M(DSP) – X – 2 (Vermont GE MX: 24) NECAP: M(DSP) – X – 3 (Vermont GE MX: 25)	4 – 39
Section 2	Uses counting techniques to solve problems... For a probability event in which the sample space...	NECAP: M(DSP) – X – 4 (Vermont GE MX: 26) NECAP: M(DSP) – X – 5 (Vermont GE MX: 27)	40 – 50

Note: Examples are provided throughout this document to illustrate definitions or phrases used in the mathematics GLEs. However, the kinds of questions students might be asked in instruction or on the NECAP assessment about the mathematics being illustrated are NOT necessarily limited to the *specific* examples given. It is the intent, over time, to add released NECAP items to expand the set of examples.

Please send comments, corrections, or suggestion to mathsurvey@gmavt.net.

* Grade-Level Expectations are called Grade Expectations (GEs) in Vermont.

Resource Material Prototype

Table of Contents

Definition	Section Number	Page Number	Definition Number
Analyzes the impact of outliers on the mean, median and mode	1	34	<i>DSP – 22</i>
Bar graph	1	8	<i>DSP – 6</i>
Box-and-whisker plot	1	14	<i>DSP – 10</i>
Circle graph	1	12	<i>DSP – 8</i>
Combination	2	43	<i>DSP – 26</i>
Dispersion	1	22	<i>DSP – 19</i>
Estimated line of best fit	1	21	<i>DSP – 14</i>
Evaluates samples from which the statistics were developed (bias)	1	39	<i>DSP – 24</i>
Event	2	47	<i>DSP – 31</i>
Experimental probability	2	50	<i>DSP – 33</i>
Frequency table	1	6	<i>DSP – 4</i>
Fundamental counting principle	2	46	<i>DSP – 29</i>
Histogram	1	10	<i>DSP – 7</i>
Identifies or describes representations or elements of representations that best display a given set of data or situation	1	35	<i>DSP – 23</i>
Interprets a given representation	1	24	<i>DSP – 21</i>
Line graph	1	17	<i>DSP – 12</i>
Line plot	1	7	<i>DSP – 5</i>
Mean	1	22	<i>DSP – 15</i>
Median	1	22	<i>DSP – 16</i>
Mode	1	22	<i>DSP – 17</i>
Outlier	1	23	<i>DSP – 20</i>
Permutation	2	44	<i>DSP – 27</i>
Pictograph	1	5	<i>DSP – 2</i>
Quartiles	1	16	<i>DSP – 11</i>
Range	1	22	<i>DSP – 18</i>
Representation	1	5	<i>DSP – 1</i>
Sample space	2	47	<i>DSP – 30</i>
Scatter plot	1	19	<i>DSP – 13</i>
Solves problems using a variety of counting strategies	2	41	<i>DSP – 25</i>
Stem-and-leaf plot	1	13	<i>DSP – 9</i>
Tally chart	1	6	<i>DSP – 3</i>
Theoretical probability	2	48	<i>DSP – 32</i>
Tree diagram	2	45	<i>DSP – 28</i>

Resource Material Prototype

Section 1: Data and Statistics

NECAP: M(DSP) – X – 1, M(DSP) – X – 2, M(DSP) – X – 3

Vermont: MX: 23, MX: 24, MX: 25

Definition	Page Number	Definition Number
Analyzes the impact of outliers on the mean, median and mode	34	<i>DSP – 22</i>
Bar graph	8	<i>DSP – 6</i>
Box-and-whisker plot	14	<i>DSP – 10</i>
Circle graph	12	<i>DSP – 8</i>
Dispersion	22	<i>DSP – 19</i>
Estimated line of best fit	21	<i>DSP – 14</i>
Evaluates sample from which the statistics were developed (bias)	39	<i>DSP – 24</i>
Frequency table	6	<i>DSP – 4</i>
Histogram	10	<i>DSP – 7</i>
Identifies or describes representations or elements of representations that best display a given set of data or situation	35	<i>DSP – 23</i>
Interprets a given representation	24	<i>DSP – 21</i>
Line graph	17	<i>DSP – 12</i>
Line plot	7	<i>DSP – 5</i>
Mean	22	<i>DSP – 15</i>
Median	22	<i>DSP – 16</i>
Mode	22	<i>DSP – 17</i>
Outlier	23	<i>DSP – 20</i>
Pictograph	5	<i>DSP – 2</i>
Quartiles	16	<i>DSP – 11</i>
Range	22	<i>DSP – 18</i>
Representation	5	<i>DSP – 1</i>
Scatter plot	19	<i>DSP – 13</i>
Stem-and-leaf plot	13	<i>DSP – 9</i>
Tally chart	6	<i>DSP – 3</i>

Resource Material Prototype

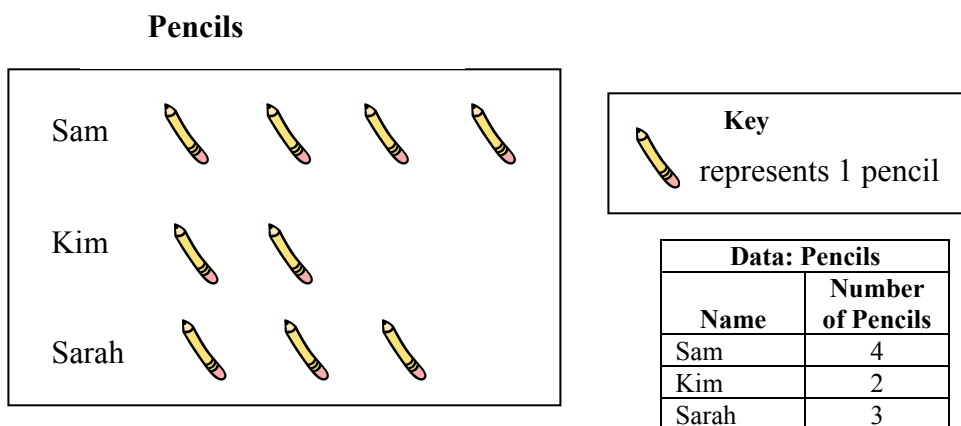
Section 1: Data and Statistics

DSP – 1 Representation: A representation refers to the type of graph, table, diagram, or model that is used to display data. The representations that are included in the K – 8 GLEs across the grades include:

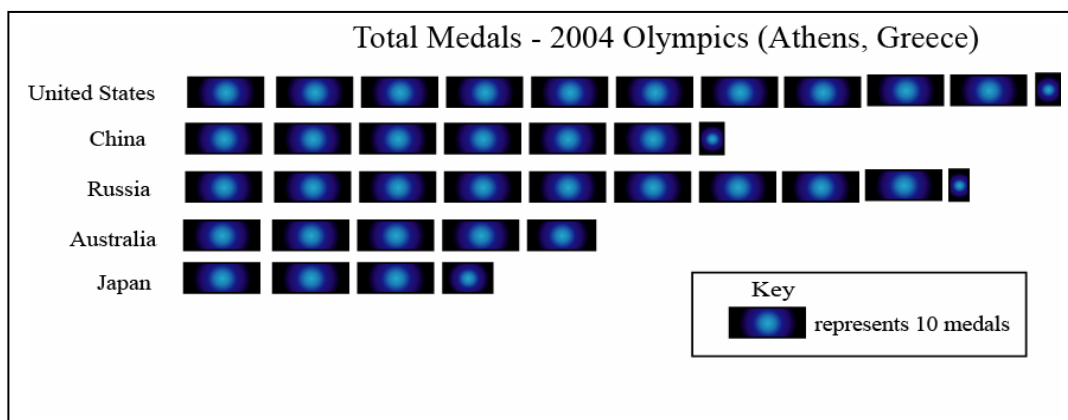
- Pictographs
- Line plots
- Tally charts
- Tables
- Bar graphs
- Circle graphs
- Line graphs
- Stem-and-leaf plots
- Scatter plots
- Histograms
- Box-and-whisker plots
- Frequency tables

DSP – 2 Pictograph: A pictograph shows numerical data using a picture or icon to represent a value.

Example 2.1 – Pictograph with one-to-one correspondence:



Example 2.2 – Pictograph where each picture or icon represents a multiple of the unit:



http://olympics.nytimes.com/results/medals_table.html

Resource Material Prototype

Section 1: Data and Statistics

DSP – 3 Tally Chart: A tally chart is a representation that consists of categories with the frequencies indicated by tally marks.

Example 3.1:

Eye Color	
Color	Number of Students
Brown	
Blue	
Green	
Hazel	

Each tally mark represents one student.

DSP – 4 Frequency table: A frequency table is a table that indicates the number of times an observation or category occurs. There are different types of frequency tables. The type described in Example 4.1 is called an absolute frequency table.

Example 4.1 – Halloween Candy:

The Number of Candy Bars Students Brought to School the Day after Halloween

Number of Candy Bars	Number of Students
0	1
1	1
2	1
3	3
4	0
5	4
6	2
7	1
8	2

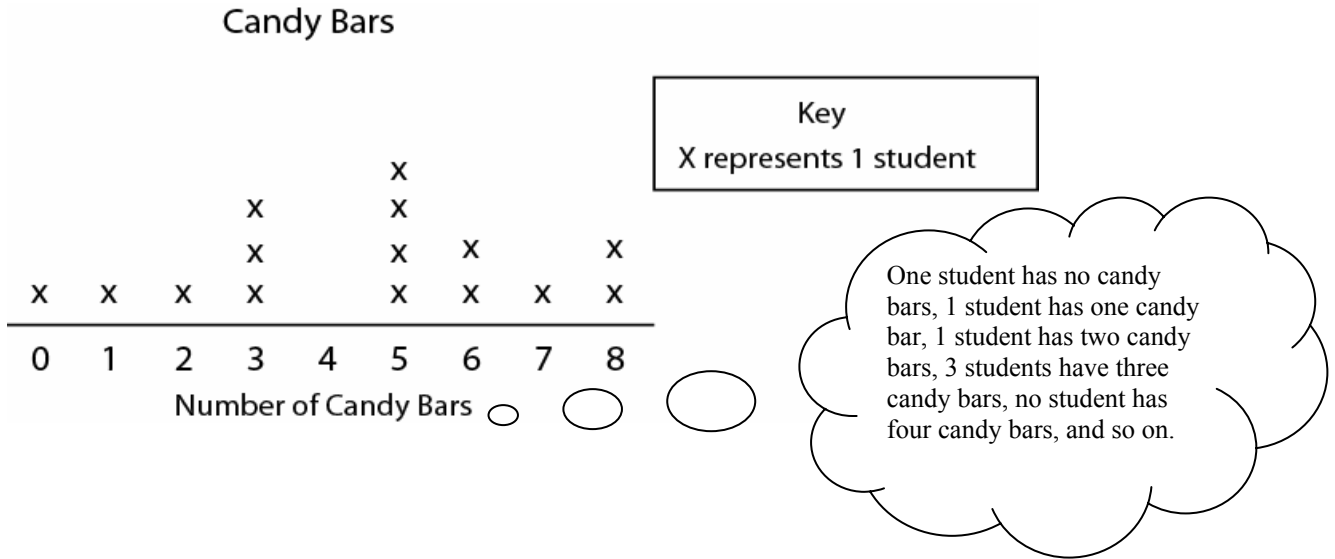
Reads as... Four students each had 5 candy bars.

Resource Material Prototype

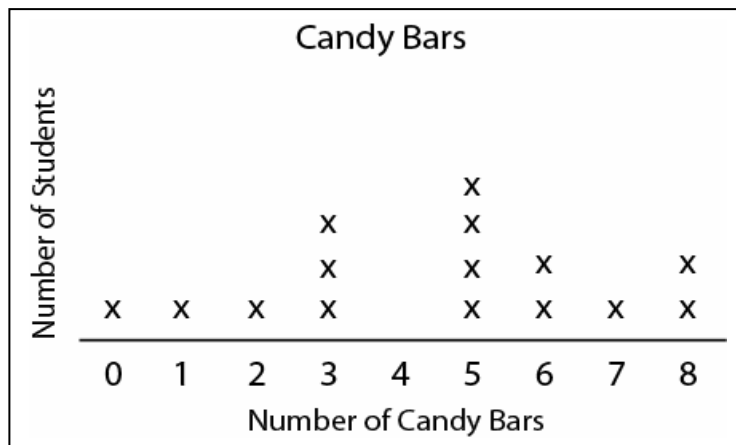
Section 1: Data and Statistics

DSP – 5 Line plot: A line plot is a representation in which X's, symbols, or objects are used to show a frequency along a number line.

Example 5.1 – Line plot with key:



Example 5.2 – Line plot with a label on the vertical axis and no key:

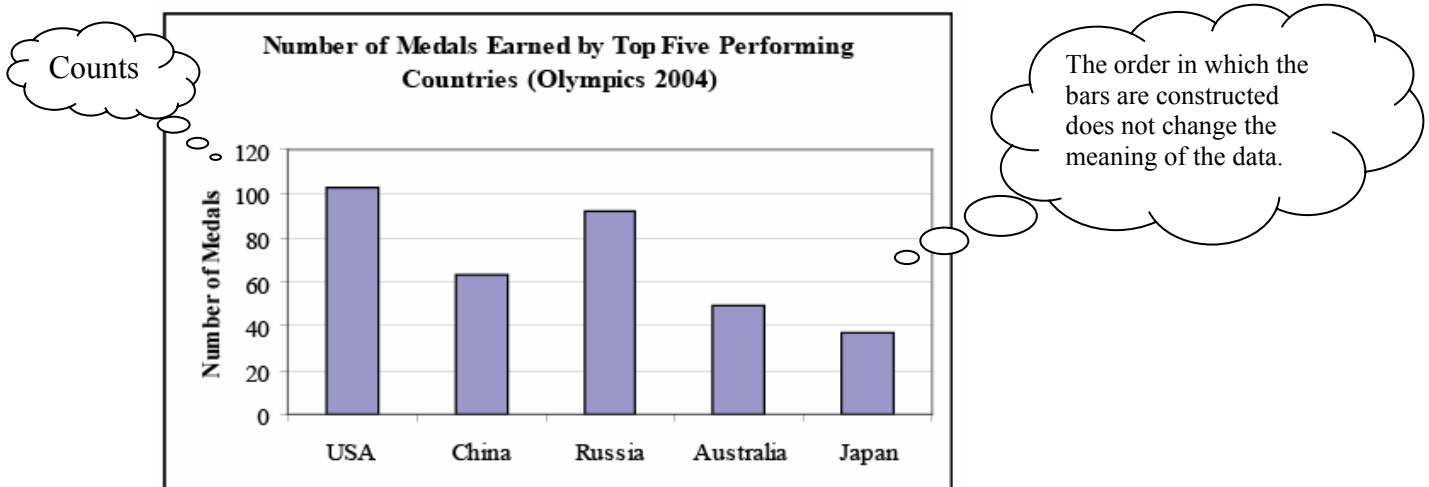


Resource Material Prototype

Section 1: Data and Statistics

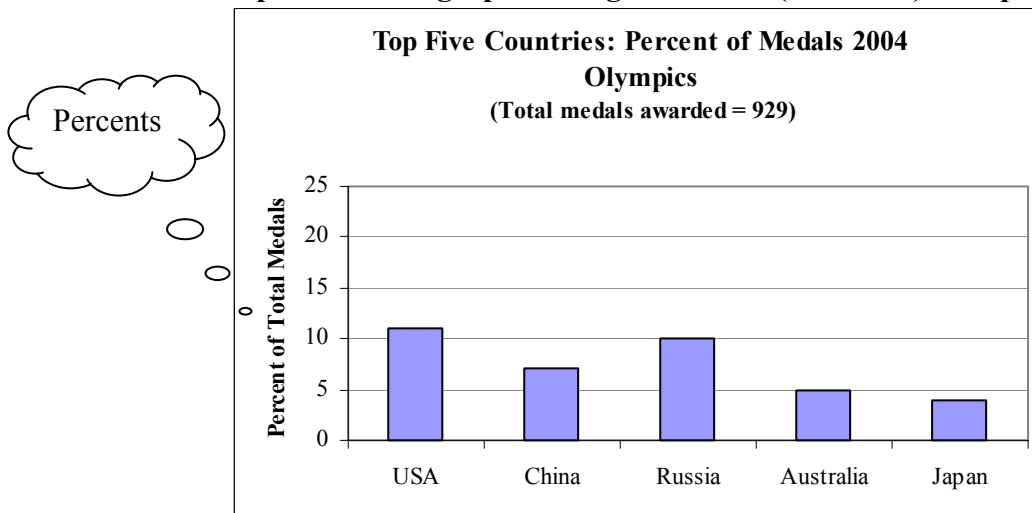
DSP – 6 Bar graph: A bar graph is a representation that is used to make comparisons among groups. Bar graphs are constructed using horizontal or vertical bars of equal widths. The heights (lengths) are equal to the values (e.g., counts, percents, temperature scales) being represented. While bar graphs are used to display both *categorical data* (e.g., countries, favorite colors, months, eye color, ages, and favorite whole number) and discrete *numerical data* it is one of the most common ways to display categorical data. An important characteristic of a bar graph displaying categorical data is that there is not always one “correct” order in which to put the categories.

Example 6.1 – Bar graph – categorical data (countries) with counts:



http://olympics.nytimes.com/results/medals_table.html

Example 6.2 – Bar graph – categorical data (countries) with percents:



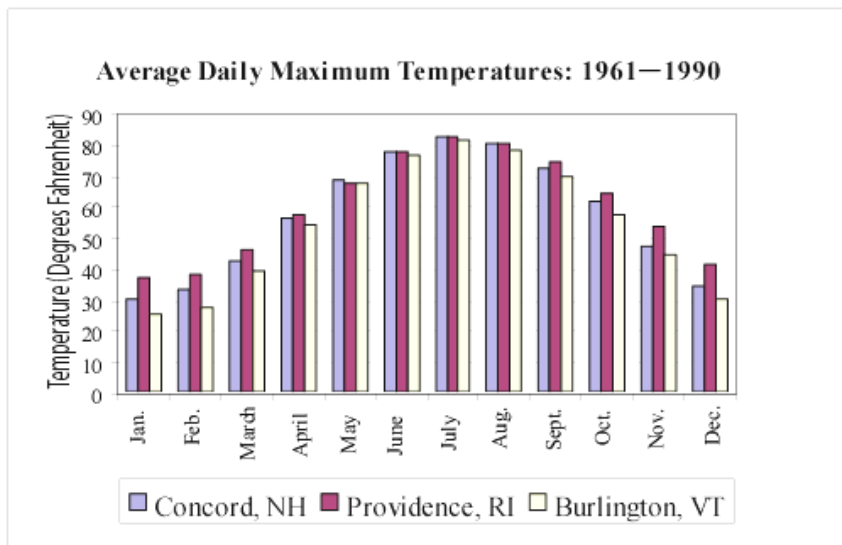
http://olympics.nytimes.com/results/medals_table.html

(Definition DSP – 6 continued on following page)

Resource Material Prototype

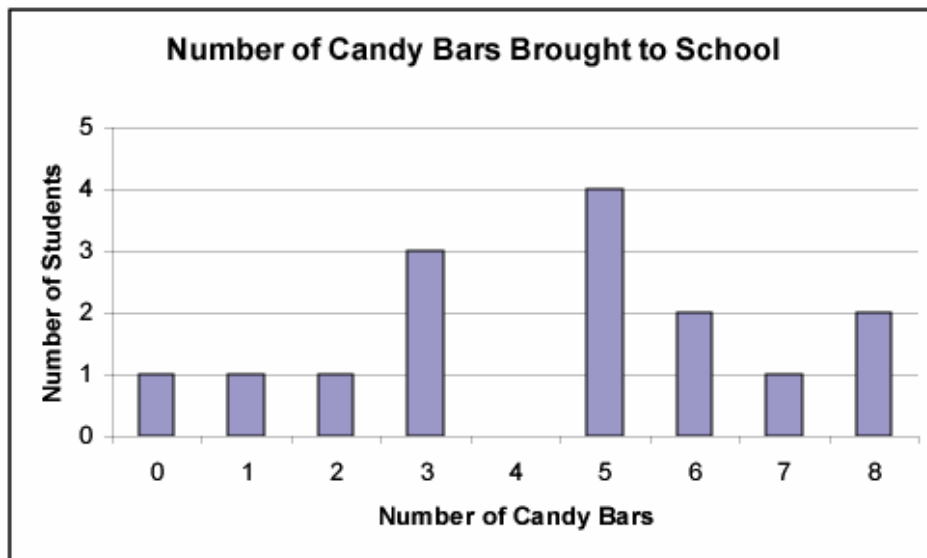
Section 1: Data and Statistics

Example 6.3 – Multiple bars to compare within and across sets of data:



<http://www.met.utah.edu/jhorel/html/wx/climate/maxtemp.html>

Example 6.4 – Bar graph representing discrete numerical data:



Resource Material Prototype

Section 1: Data and Statistics

DSP – 7 Histogram: A histogram is a graphical representation of frequency distributions (e.g., counts, percents) used to compare groups in which the groups have values that are continuous and can be ordered from least to greatest. Some variables that can be represented using histograms include ages, heights, test scores, and times.

A histogram is constructed of contiguous rectangles (rectangles touching each other) with:

- widths of the rectangles representing numerical intervals (groups) that are determined by the person who organizes the histogram (See examples 7.2 and 7.3.);
- widths of the rectangles proportional to the size of the respective intervals (e.g., If all the numerical intervals are the same length, then the widths of the rectangles are the same length. However, if, for example, one interval spans 10 years and another interval spans 20 years, then the width of the rectangle representing 20 years is twice as long as the width of the rectangle representing 10 years.); and
- areas of the rectangles proportional to the percent or frequency (height of rectangle).

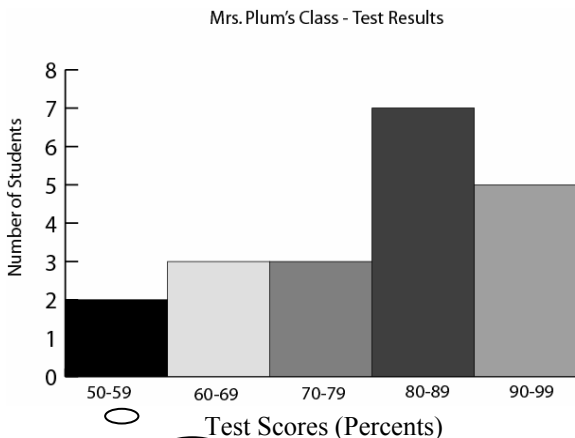
NECAP will *only* assess understanding of histograms with class intervals of equal widths.**

Mrs. Plum's Class

Name	Percent*
Barbara	94
Beth	61
Bob	62
Brian	84
Chris	80
Frank	93
Hannah	84
Ian	88
Karen	84
Kim	51
Marge	72
Nick	98
Patty	96
Phil	71
Ralph	53
Richard	89
Sam	60
Sam R.	87
Sonia	97
Ted	72

*Percents rounded to the nearest whole percent

Example 7.1 – Histogram with class intervals equal in size (10 percentage points):



Test Score Intervals	Frequency
50 – 59	2
60 – 69	3
70 – 79	3
80 – 89	7
90 – 99	5

The area of each rectangle is proportional to the frequency of its respective class. (i.e., The area of each rectangle is ten times the frequency.)

** For information about histograms with unequal intervals see Freedman, D., Pisani, R., and Purves, R., (1997) *Statistics, 3rd. Edition*. W.W. Norton and Company.

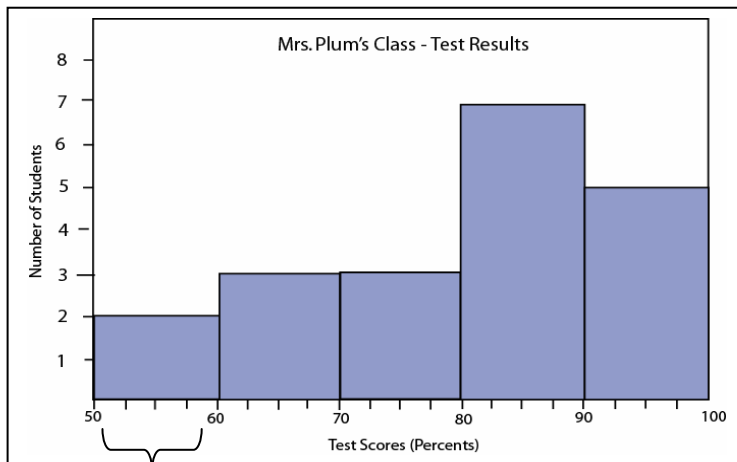
(Definition DSP – 7 continued on following page)

Resource Material Prototype

Section 1: Data and Statistics

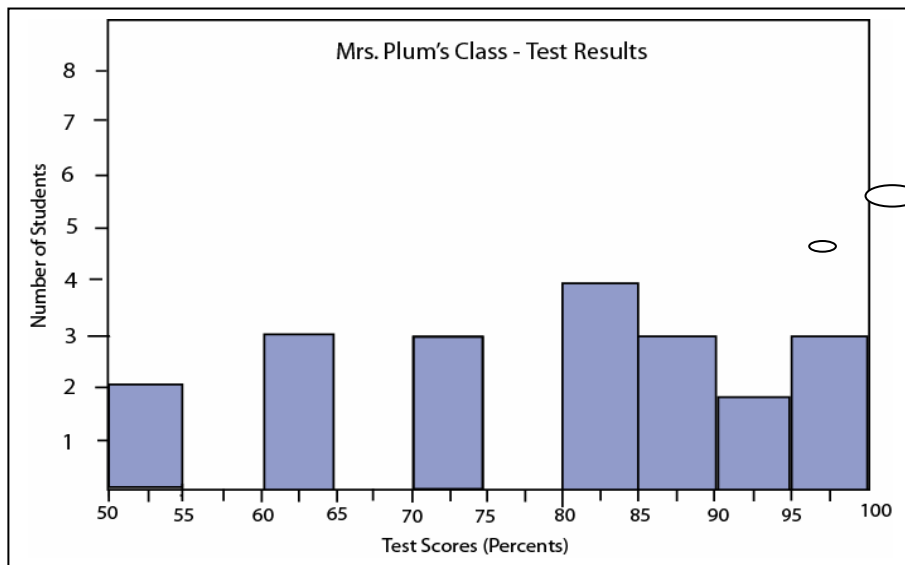
Example 7.2 – Histogram with class intervals equal in size (10 percent points) with intervals on a scale:

Test Score Intervals	Frequency
50–59	2
60–69	3
70–79	3
80–89	7
90–99	5



When interpreting histograms with a scale the typical convention is to include the left endpoint in the interval and exclude the right endpoint (e.g., Test scores 50, 51, 52, 53, 54, 55, 56, 57, 58, and 59 are included in the first bar in the graph, but not test score 60. Test score 60 is included in the second bar of the

Example 7.3 – Histogram with class intervals equal in size (5 percentage points) with intervals on a scale:



Whoever organizes the histogram decides the length of the intervals.

Test Score Intervals	Frequency
50 – 54	2
55 – 59	0
60 – 64	3
65 – 69	0
70 – 74	3
75 – 79	0
80 – 84	4
85 – 89	3
90 – 94	2
95 – 99	3

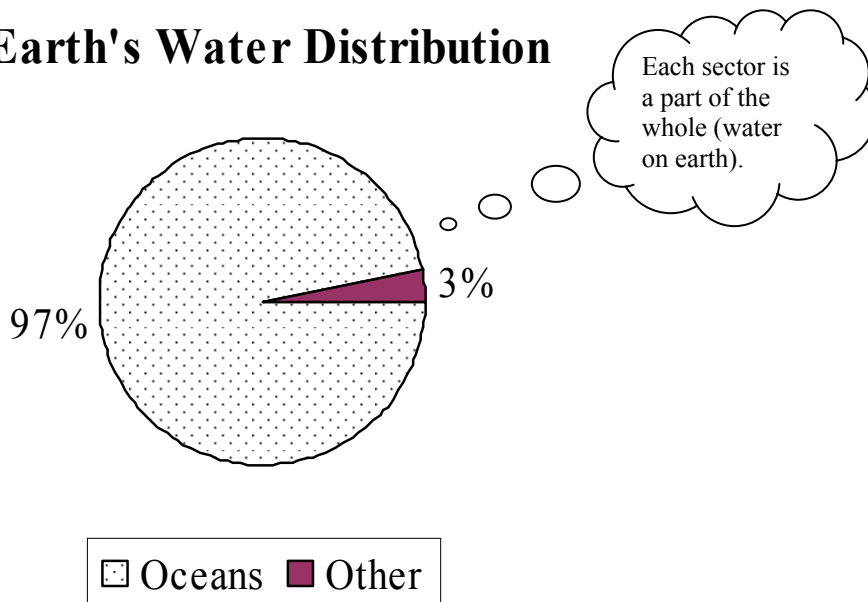
Resource Material Prototype

Section 1: Data and Statistics

DSP – 8 Circle graph: A circle graph is a pictorial representation used to compare parts of a whole. The area of the circle represents the whole and is divided into pieces called sectors that represent parts of a whole. The size of the sector is determined by the percent of the whole that its respective categories represent.

Example 8.1 – Circle graph:

Earth's Water Distribution



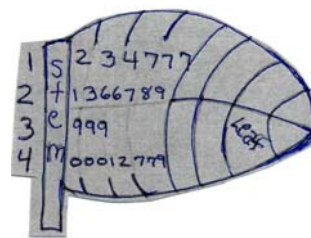
<http://ga.water.usgs.gov/edu/waterdistribution>

Resource Material Prototype

Section 1: Data and Statistics

DSP – 9 Stem-and-leaf plot: A stem-and-leaf plot organizes data to show their shape and distribution. Each data value is split into a “stem” and a “leaf.” Typically, the stem is the first digit(s) of a data value and the leaf is the last digit of the data value.

(e.g., If 27 is a data value, then 2 is the “stem” and 7 is the “leaf.”)



Example 9.1 – Mathematics Test Scores– Mrs. Plum and Mr. Scarlet:

Mr. Scarlet and Mrs. Plum gave the same test to two different classes. The results from the tests are organized into the tables and stem-and-leaf plots below.

e.g., 5 | 1
represents 51%

Mrs. Plum's Class	
5	1, 3
6	0, 1, 2
7	1, 2, 2
8	0, 4, 4, 4, 7, 8, 9
9	3, 4, 6, 7, 8

Stems

Mr. Scarlet's Class	
6	5, 8, 8, 8
7	0, 0, 0, 6, 6, 6, 8, 9
8	0, 1, 2, 5, 5, 8, 8
9	7, 8

Leaves

Stem-and-leaf plots concisely organize large amounts of data in a way that reveals the distribution of the data, the mode, and provides an easy way to determine the median of the data set.

Mrs. Plum's class

Name	Percent*
Barbara	94
Beth	61
Bob	62
Brian	84
Chris	80
Frank	93
Hannah	84
Ian	88
Karen	84
Kim	51
Marge	72
Nick	98
Patty	96
Phil	71
Ralph	53
Richard	89
Sam	60
Sam R.	87
Sonia	97
Ted	72

*Percents rounded to the nearest whole percent

Mr. Scarlet's class

Name	Percent*
Beth	97
Bob	76
Brian	78
Chris	65
Darlene	76
Debbie	70
Glenn	70
Ian	85
Kim	68
Lee	70
Leslie	80
Lorie	79
Lucinda	81
Marilyn	88
Mike	68
Mike R.	76
Nick	85
Ralph	98
Richard	82
Ted	68
Tom	88

*Percents rounded to the nearest whole percent

Resource Material Prototype

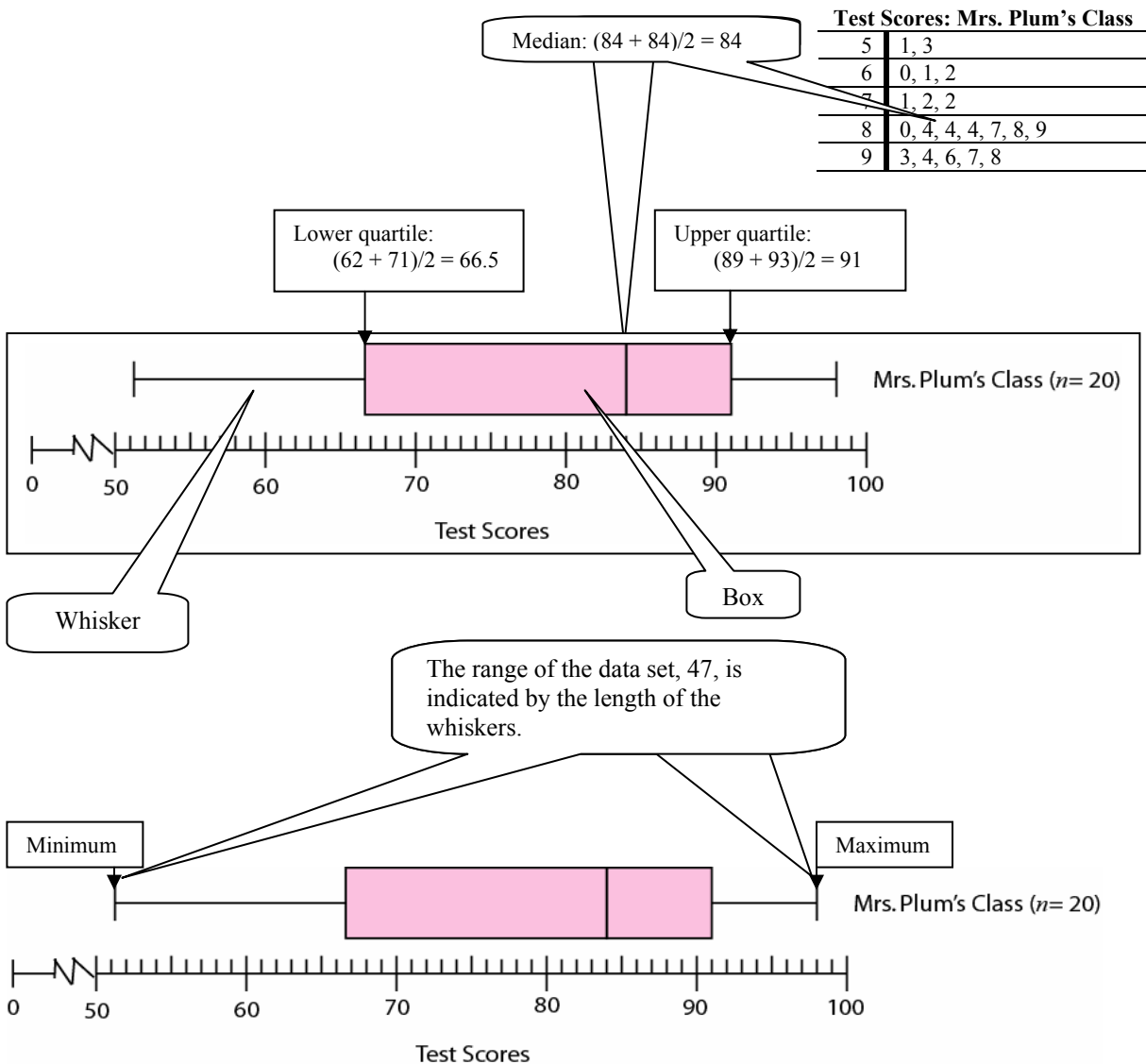
Section 1: Data and Statistics

DSP – 10 Box-and-whisker plot: A box-and-whisker plot is a representation which visually displays the median of a data set, the upper and lower quartiles, and information about the range and distribution of the data.

To construct a box-and-whisker plot:

- 1) Find the values of the median, and the upper and lower quartiles.
- 2) Draw a rectangle (called the box) from the lower quartile to the upper quartile, and draw a line across the width of the box at the median.
- 3) Extend a line (called the whiskers) from the midpoint of the box to the minimum value and similarly draw a line extending to the maximum value.

Example 10.1 – Box-and-whisker plot (Data from Example 9.1):



(Definition DSP – 10 continued on following page)

Resource Material Prototype

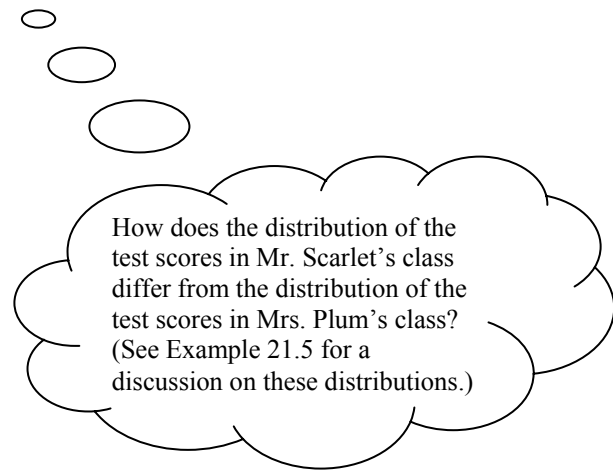
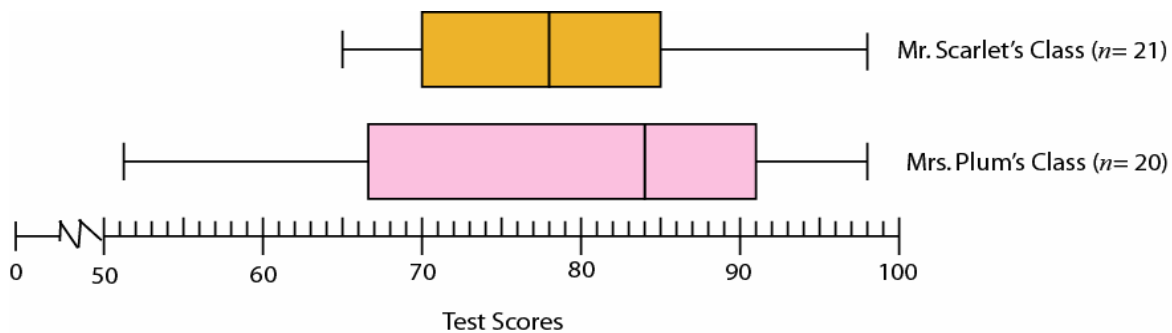
Section 1: Data and Statistics

Example 10.2 – Box-and-whisker plots are effective representations to compare distributions between sets of data as in the test score data below:

Mathematics Test Scores: Mrs. Plum’s class and Mr. Scarlet’s class

Mrs. Plum’s Class	
5	1, 3
6	0, 1, 2
7	1, 2, 2
8	0, 4, 4, 4, 7, 8, 9
9	3, 4, 6, 7, 8

Mr. Scarlet’s Class	
6	5, 8, 8, 8
7	0, 0, 0, 6, 6, 6, 8, 9
8	0, 1, 2, 5, 5, 8, 8
9	7, 8

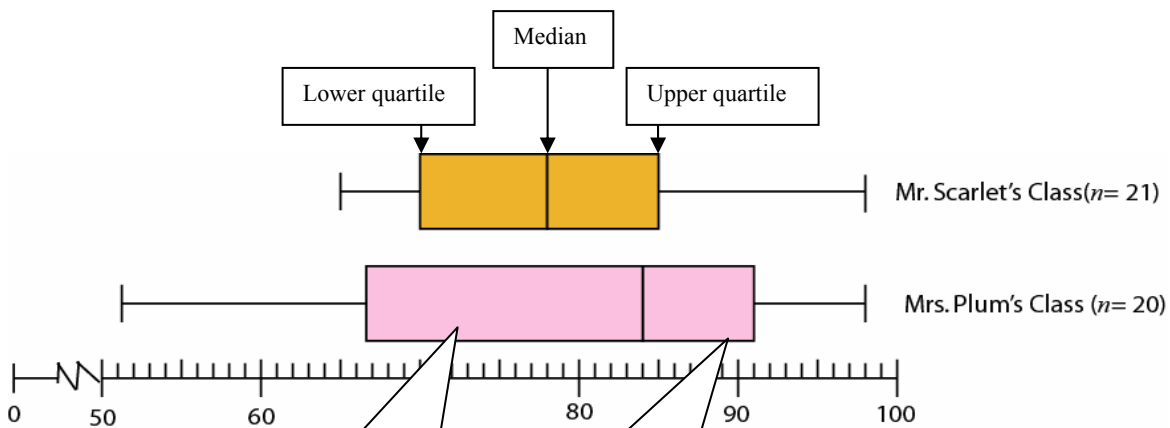


Resource Material Prototype

Section 1: Data and Statistics

DSP – 11 Quartiles: Quartiles are values that divide a set of data (arranged in numerical order) into four groups. The points that divide these groups are called the lower quartile, middle quartile (or median), and the upper quartile (Note: There are other ways to name these such as first, second, and third quartile.). Each group contains about 25% of the data points in the set of data depending upon the size and nature (e.g., odd number of data points, even number of data points) of the data set (e.g., about 75% of the data values are at or above the lower quartile.).

Example 11.1 – Box plots are a way to represent data organized by quartiles:



The box between the lower quartile and the median is longer than the box between the median and the upper quartile. This means that the spread of the data between the lower quartile and the median is greater than the spread of the data between the median and the upper quartile.

Test Scores: Mrs. Plum's Class

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

Box-and-whisker plots do not indicate how the data are spread between quartile values (e.g., while the range of the data values in Mrs. Plum's class between the lower quartile and the median is from 66.5% to 84%, there are no scores between 73% and 79%).

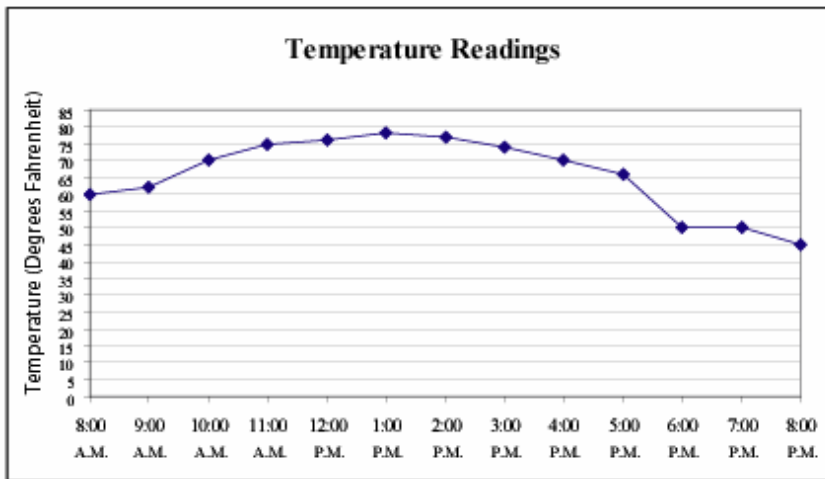
Resource Material Prototype

Section 1: Data and Statistics

DSP – 12 Line graph: A line graph is typically used to represent serial data (data that can be collected without gaps such as change in temperature over a time period).

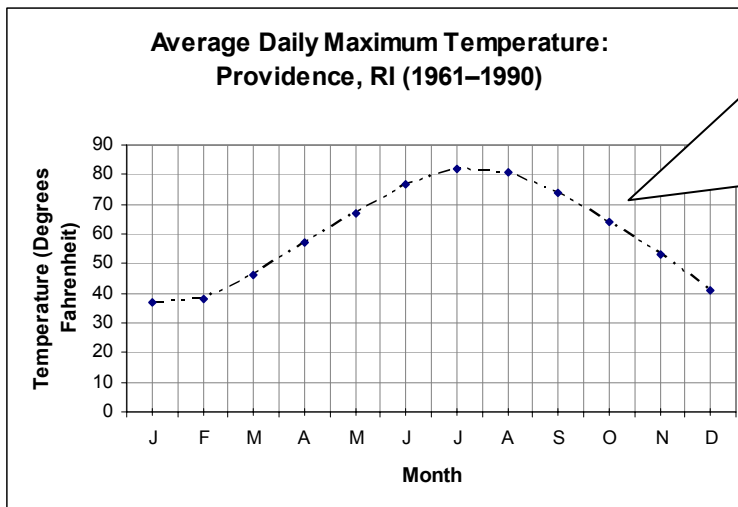
Typically, observations are made over regular intervals. Then, these discrete points are plotted and adjacent points are connected with line segments to indicate the serial nature of the data and allow for interpolation.

Example 12.1 – Line Graph:



Time of Day	Temperature (°F)
8:00 A.M.	60
9:00 A.M.	62
10:00 A.M.	70
11:00 A.M.	75
12:00 P.M.	76
1:00 P.M.	78
2:00 P.M.	77
3:00 P.M.	74
4:00 P.M.	70
5:00 P.M.	66
6:00 P.M.	50
7:00 P.M.	50
8:00 P.M.	45

Example 12.2 – Broken Line Graph:



A broken line graph can be used to show the trends in non-serial data, such as the average daily maximum temperatures from month to month. The line is broken because the data points are NOT serial as in Example 12.1. That is, drawing a line between any two points would imply that an average daily maximum temperature exists, for example, for some month between January and February.

J	F	M	A	M	J	J	A	S	O	N	D
37	38	46	57	67	77	82	81	74	64	53	41

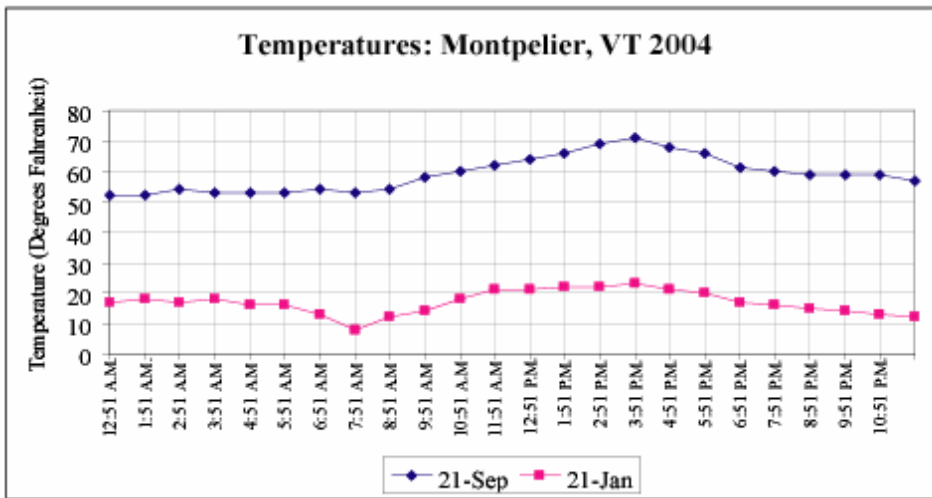
<http://www.met.utah.edu/jhorel/html/wx/climate/maxtemp.html>

(Definition DSP – 12 continued on following page)

Resource Material Prototype

Section 1: Data and Statistics

Example 12.3 – Double line graph:



Montpelier, VT 2004		
	21-Sep	21-Jan
12:51 A.M.	52	17
1:51 A.M.	52	18
2:51 A.M.	54	17
3:51 A.M.	53	18
4:51 A.M.	53	16
5:51 A.M.	53	16
6:51 A.M.	54	13
7:51 A.M.	53	8
8:51 A.M.	54	12
9:51 A.M.	58	14
10:51 A.M.	60	18
11:51 A.M.	62	21
12:51 P.M.	64	21
1:51 P.M.	66	22
2:51 P.M.	69	22
3:51 P.M.	71	23
4:51 P.M.	68	21
5:51 P.M.	66	20
6:51 P.M.	61	17
7:51 P.M.	60	16
8:51 P.M.	59	15
9:51 P.M.	59	14
10:51 P.M.	59	13
11:51 P.M.	57	12

<http://www.wunderground.com/US/VT/Montpelier.html>

Resource Material Prototype

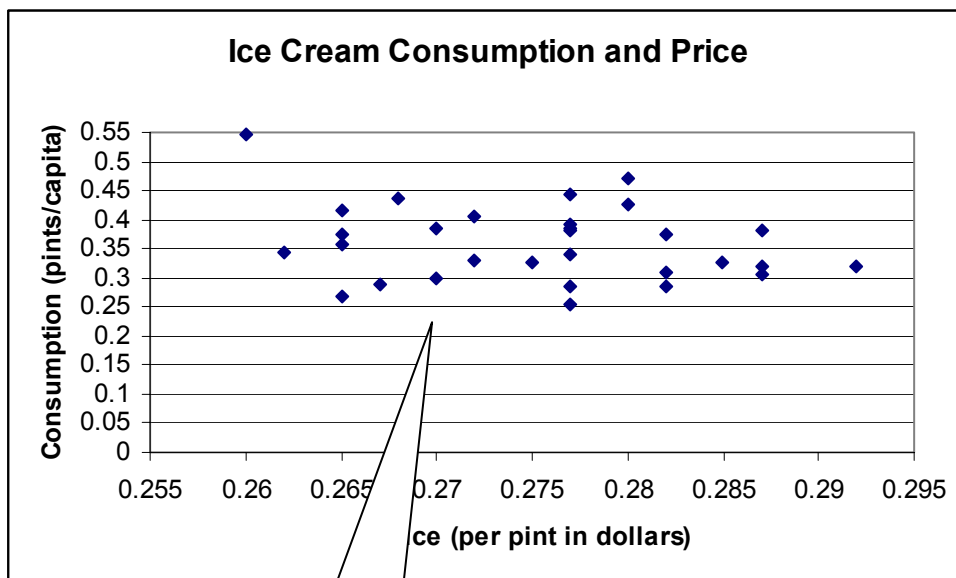
Section 1: Data and Statistics

DSP – 13 Scatter plot: Scatter plots are plots of observations that are used to determine if there is a relationship between two variables being studied. Scatter plots help to answer questions about how two variables are related to each other (e.g., Is there a relationship between the amount of time students watch TV and their grades?).

Examples 13.1 and 13.2 contain data collected over 30 four-week periods from March 18, 1951 to July 11, 1953.

Example 13.1 – Relationship of Ice Cream Consumed to Price of Ice Cream:

The question being asked in this example was “Is ice cream consumption dependent upon price?”



There does not appear to be any discernable trend in the data.

Ice Cream Consumption (Pints consumed per capita)	Price (per pint in dollars)	Temperature (degrees Fahrenheit)
0.386	0.270	41
0.374	0.282	56
0.393	0.277	63
0.425	0.280	68
0.406	0.272	69
0.344	0.262	65
0.327	0.275	61
0.288	0.267	47
0.269	0.265	32
0.256	0.277	24
0.286	0.282	26
0.298	0.270	26
0.329	0.272	32
0.318	0.287	40
0.381	0.277	55
0.381	0.287	63
0.470	0.280	72
0.443	0.277	72
0.386	0.277	67
0.342	0.277	60
0.319	0.292	44
0.307	0.287	40
0.284	0.277	32
0.326	0.285	27
0.309	0.282	28
0.359	0.265	33
0.376	0.265	41
0.416	0.265	52
0.437	0.268	64
0.548	0.260	71

(Definition DSP – 13 continued on following page)

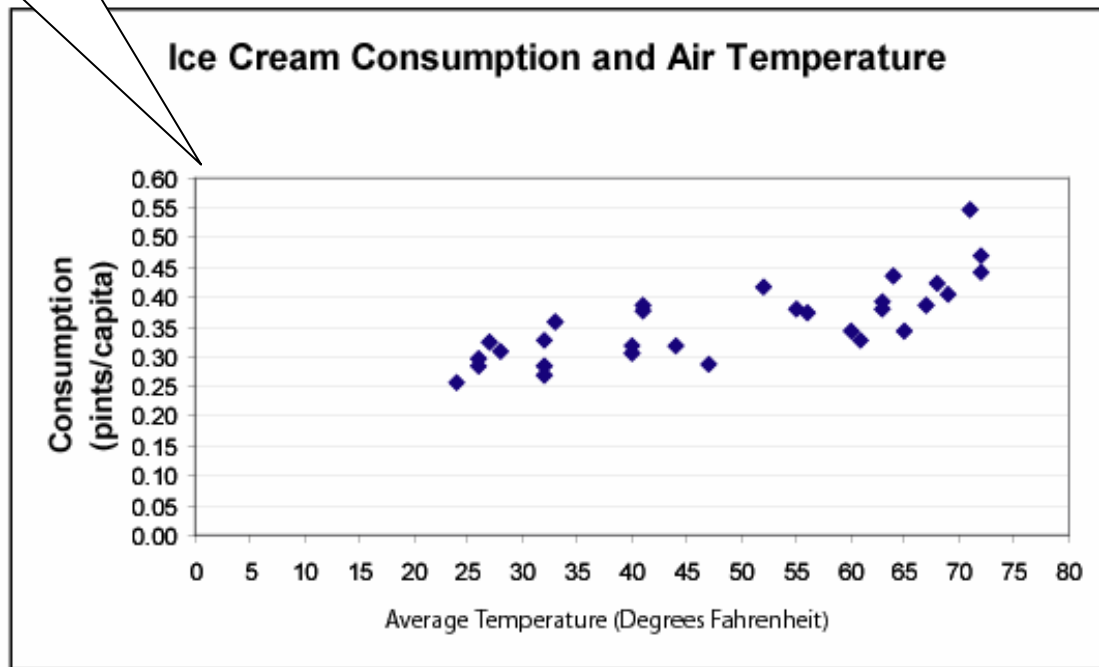
Resource Material Prototype

Section 1: Data and Statistics

Example 13.2 – Relationship of Ice Cream Consumed to Air Temperature:

The question being asked in this example was “Is there a relationship between air temperature and the amount of ice cream consumed?”

In general, as the temperature increases from 24°F to 72°F, so does ice cream consumption.



Note: Additional studies were conducted to determine if other variables have an impact on ice cream consumption, including Lag-Temp (Temp lagged by one time period) and Year, which were both significant predictors of ice cream consumption. For more information about the study go to <http://lib.stat.cmu.edu/DASL/Stories/IceCreamConsumption.html>.

Source: Koteswra Rao Kadiyala (1970). Testing for the independence of regression disturbances. *Econometrica*, 38, 97–117. Also found in: Hand, D.J., et al (1994). *A handbook of small data sets*.

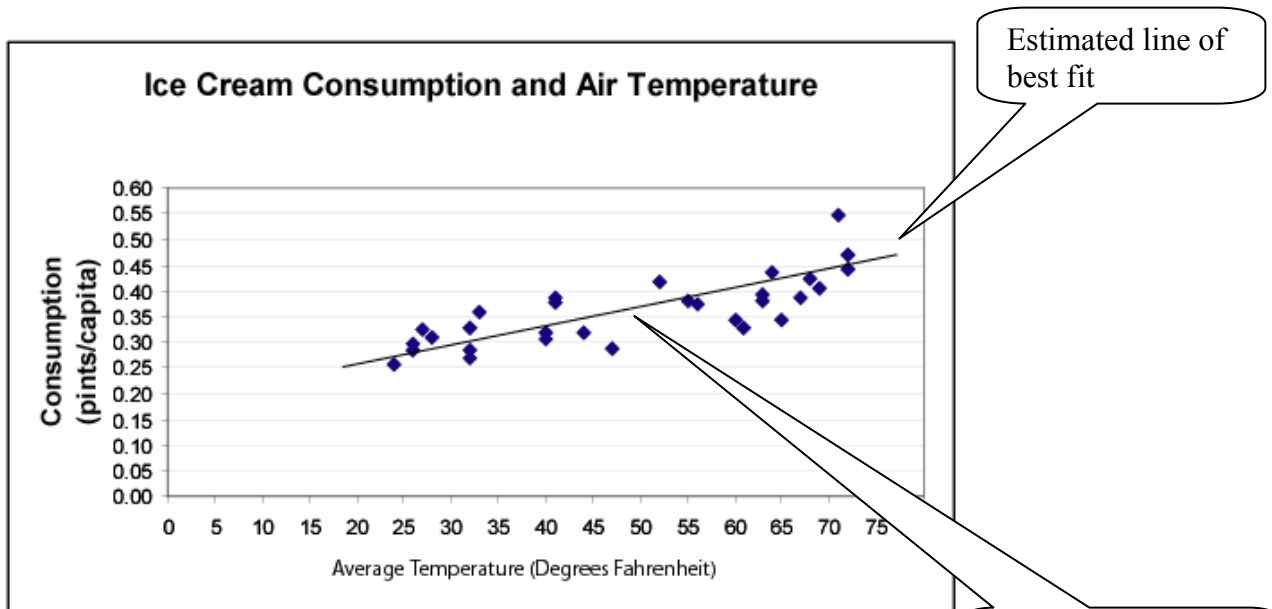
Resource Material Prototype

Section 1: Data and Statistics

DSP – 14 Estimated line of best fit: A line that can be drawn through the data points on a scatter plot that represents the “best approximation” of the trend reflected in the data¹. The line can then be used to make predictions.

Example 14.1 – Estimated line of best fit (Data from Examples 13.2):

An ice cream company would be very interested in the data from the study in Example 13.2. However, they would not just be interested in the fact that there is a trend, but they would want to use the data to determine how much ice cream to produce at different times of the year for different regions of the country (or world) based upon regional temperatures². To make a prediction about production, find an estimated line of best fit and use it to make a prediction by interpolating the values along the line at a given point.



<http://lib.stat.cmu.edu/DASL/Stories/IceCreamConsumption.htm> .

¹ The actual calculation for a line of best fit is beyond the scope of the K – 8 GLEs. A property of a line of best fit known as the least squares regression line is that the sum of the errors (difference between the actual points and the points on the line for given x -values) is zero. This means that half of the error is below the line and half of the error is above the line. Estimating this line leads to an appropriate estimated line of best fit.

² An ice cream company would have collected many more data points over longer periods of time than were collected in this study to make this type of decision. However, this example is used for illustrative purposes.

Resource Material Prototype

Section 1: Data and Statistics

DSP – 15 Mean: The mean of a data set is the sum of all the values divided by the number of values.

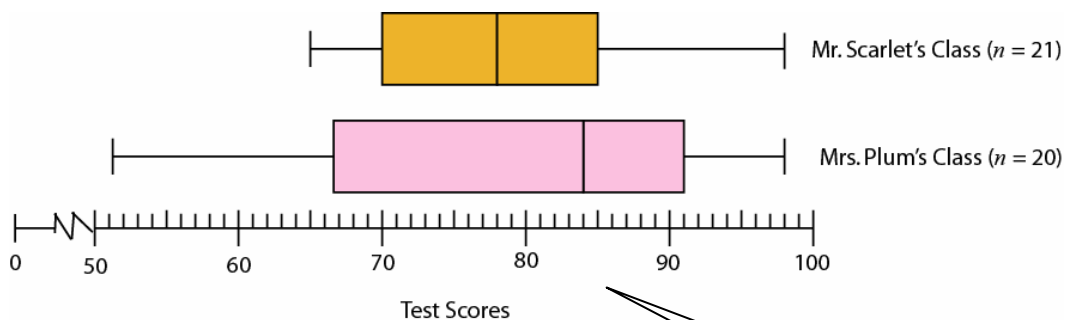
DSP – 16 Median: The median of a data set is the middle value or average of the middle values when the values are arranged in numerical order.

DSP – 17 Mode: The mode of a data set is the value that occurs most often.

DSP – 18 Range: The range is the difference between the greatest value in a data set and the least value in a data set.

DSP – 19 Dispersion: Dispersion considers the distribution of data around an average (mean, median) or between quartiles across the range of a set of data.

Example 19.1 – Box plots are one way to represent the dispersion of data around the median and between quartiles:



Box-and-whisker plots include the range of the data, the median, and how the data is spread between quartiles.

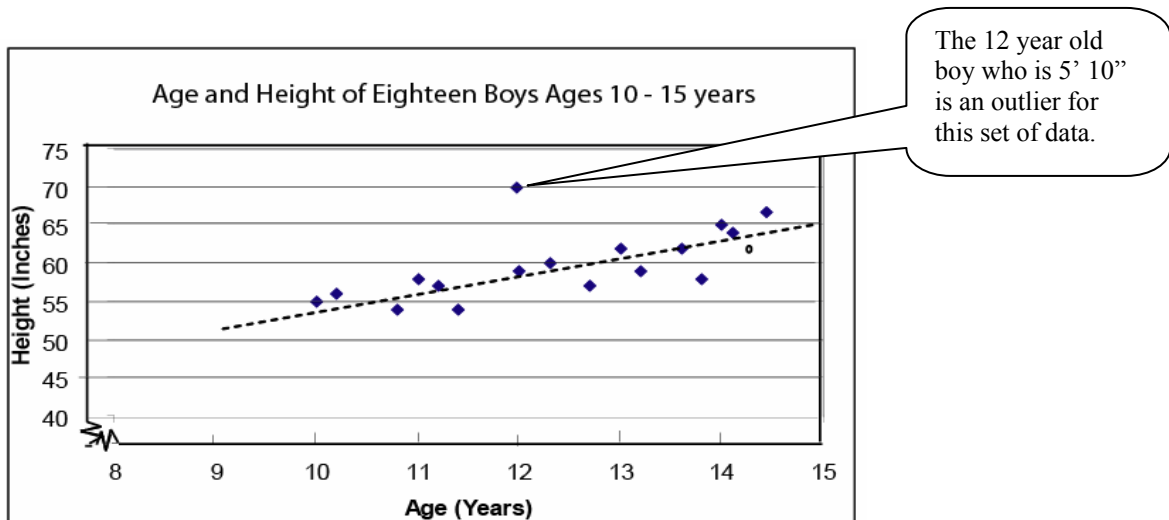
Note: Recall, box-and-whisker plots do NOT indicate how the data are spread between quartile values. Therefore, one needs to be careful if trying to use a box-and-whisker plot to determine the SHAPE of a distribution (e.g., two sets of data may generate the same box-and-whisker plot but have very different shapes).

Resource Material Prototype

Section 1: Data and Statistics

DSP – 20 Outlier: An outlier is a data point that is widely separated from the general distribution of a set of data.

Example 20.1 – Scatter plot showing the general distribution of a set of data with an outlier:



Data set constructed based on National Center for Health Statistics Growth Chart
<http://www.cdc.gov/nchs/data/nhanes/growthcharts/set2/chart%2007.pdf>.

Resource Material Prototype

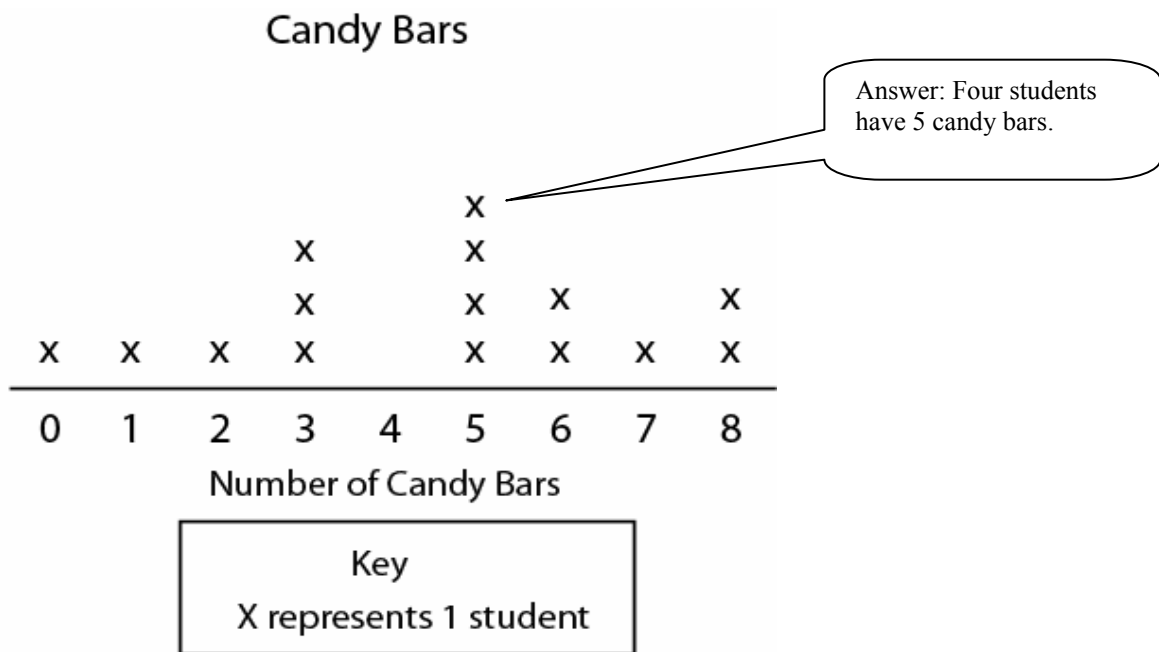
Section 1: Data and Statistics

DSP – 21 Interprets a given representation: This stem focuses on students interpreting representations of data consistent with the expectations at given grade levels. This includes answering questions about the data, analyzing data to draw and justify conclusions, making predictions based upon the data, or solving problems using the data and representations. This GLE, M(DSP)–X–1, interacts with M(DSP)–X–2. That means that when asking a student to analyze data the expectations in M(DSP)–X–2 should be considered (See Example 21.7).

Examples 21.1 – 21.10 illustrate what it means to “answer questions related to the data, to analyze the data to formulate and justify conclusions, to make predictions based on the data, or to solve problems.”

Example 21.1 – Answer questions related to the data:

How many students have 5 candy bars?



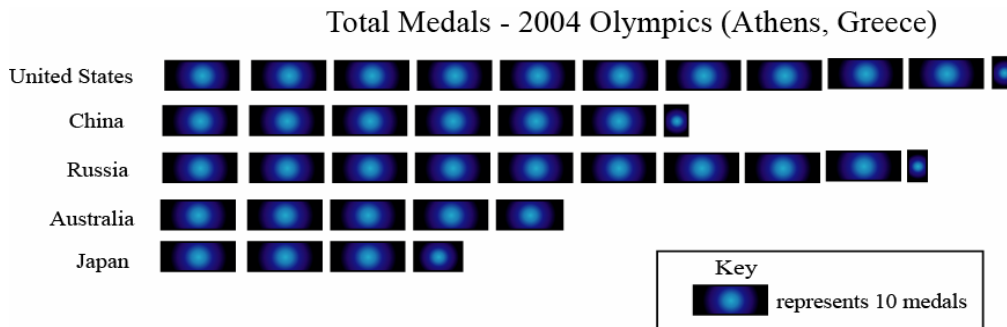
(Definition *DSP – 21* continued on following page)

Resource Material Prototype

Section 1: Data and Statistics

Example 21.2 – Answer questions related to the data:

Look at this pictograph.



Based on these data, *about* how many more medals did the United States win than Australia?

Answer: From the pictograph, it can be seen that the United States has about $5 + \frac{4}{10}$ more rectangles than Australia (counting the last rectangle for Australia as $\frac{9}{10}$). Since each rectangle represents 10 medals, the United States won about $50 + 4$ or 54 more medals.

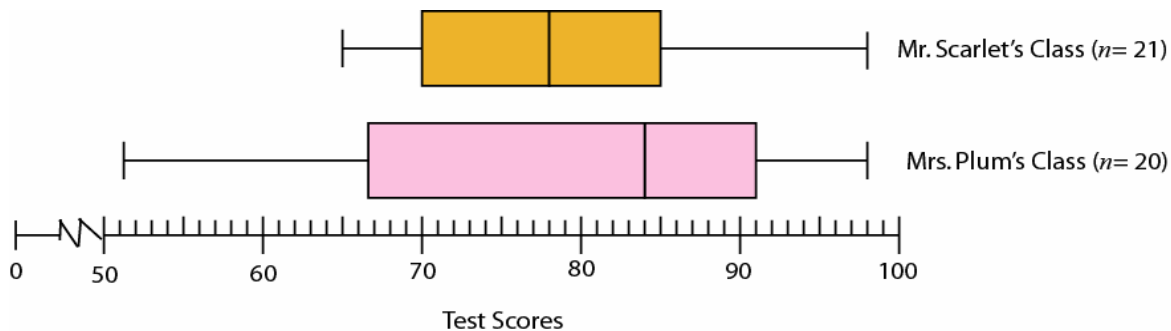
(Definition *DSP* – 21 continued on following page)

Resource Material Prototype

Section 1: Data and Statistics

Example 21.3 – Answer questions related to the data:

How much greater is the median score for Mrs. Plum’s class than the median score for Mr. Scarlet’s class?

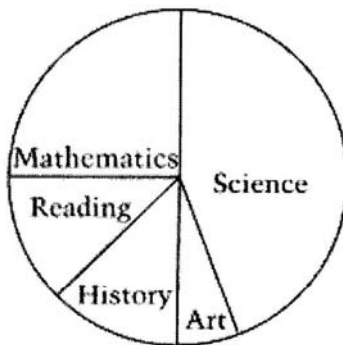


Answer: 6%; The median score for Mr. Scarlet’s class is 78%. The median for Mrs. Plum’s class is 84%. The median score for Mrs. Plum’s class is 6% greater than the median score for Mr. Scarlet’s class.

Example 21.4 – Answer questions related to the data:

2003 NAEP Released Item

Portion of Time Pat Spent on Homework



3. The pie chart above shows the portion of time Pat spent on homework in each subject last week. If Pat spent 2 hours on mathematics, about how many hours did Pat spend on homework altogether?

- A) 4
- B) 8
- C) 12
- D) 16

Answer: B

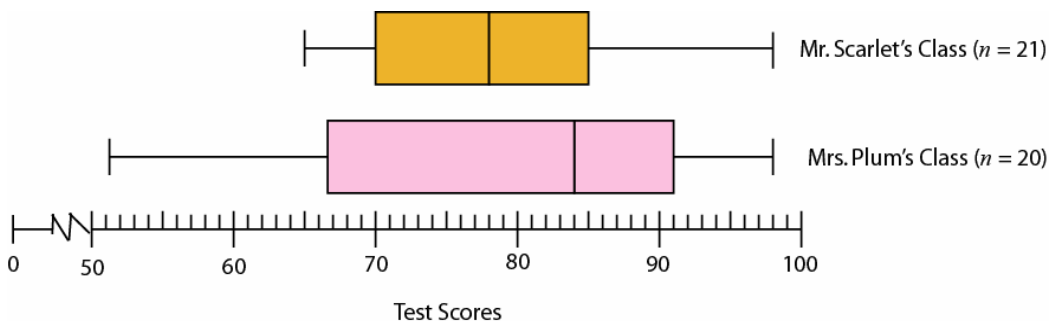
(Definition *DSP* – 21 continued on following page)

Resource Material Prototype

Section 1: Data and Statistics

Example 21.5 – Answers questions related to the data:

The following statements were made about the test results of Mr. Scarlet's class and Mrs. Plum's class using the box-and-whisker plots below. Indicate whether each statement is true or false or can not be determined, and explain why using information displayed in the box-and-whisker plots.



- A) More students in Mrs. Plum's class scored between 66.5% and 84% than between 84% AND 91% because the box is longer between 66.5% and 84% than between 84% AND 91%.
- B) The range of scores in both classes is the same.
- C) The median score in Mrs. Plum's class is higher than the median score in Mr. Scarlet's class.
- D) Because about 50% of the scores in Mr. Scarlet's class are above 78 and about 50% of the scores in Mrs. Plum class are above 84, the mean score for Mrs. Plum's class must be higher than the mean score for Mr. Scarlet's class.

Answer: Using ONLY the information represented in the box plot...

- A. This statement is false. The length of the box is determined by the spread within a quartile, not by the number of data points. Furthermore, the number of data points (student test scores) between each quartile is about 25% of the number of data points in the set of data.
- B. This statement is false. The range (difference between highest and lowest score) is greater in Mrs. Plum's class than in Mr. Scarlet's class. The fact that the distances from one end of the whisker to the other in the two box plots being compared are not the same indicates that the range is different.
- C. This statement is true. The median score in Mr. Scarlet's class is 78 while the median score in Mrs. Plum's class is 84.
- D. This statement is false. You can not make a statement about the mean score using the distribution represented in a box plot. You would need to know more about how the data is distributed between the minimum score and the lower quartile, the lower quartile and the median, and so on. For example, the (approximately) 50% of the scores below 78 in Mr. Scarlet's class may be closely clumped around 77 and 70, whereas the 50% of the scores below 84 in Mrs. Plum's class may all be clumped around 67 and 51, causing the mean score in Mr. Scarlet's class to be greater than the mean score in Mr. Plum's class.

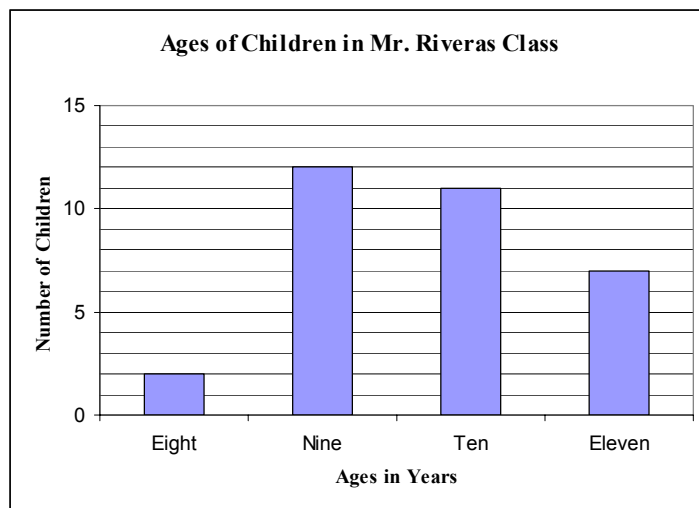
(Definition DSP – 21 continued on following page)

Resource Material Prototype

Section 1: Data and Statistics

Example 21.6 – Answer questions related to the data:

Adapted 1992 NAEP
Released Item



The graph above shows how many of the 32 children in Mr. Rivera's class are eight, nine, ten, and eleven years old. Which of the following is true?

- A) Most are younger than nine.
- B) Most are younger than ten.
- C) Most are nine or older.
- D) Most are older than eleven.

Answer: C

(Definition *DSP – 21* continued on following page)

Resource Material Prototype

Section 1: Data and Statistics

Example 21.7 – Interprets a given representation to analyze the data to formulate or justify conclusions AND Analyzes patterns, trends, or distributions in a variety of contexts by solving problems using measures of central tendency...:

This example illustrates how $M(DSP) - X - 1$ interacts with $M(DSP) - X - 2$.

Mr. Scarlet and Mrs. Plum gave the same mathematics test to each of their classes. They displayed the results from the test into the histograms and box-and-whisker plots shown below, and determined the mean and mode for each class.

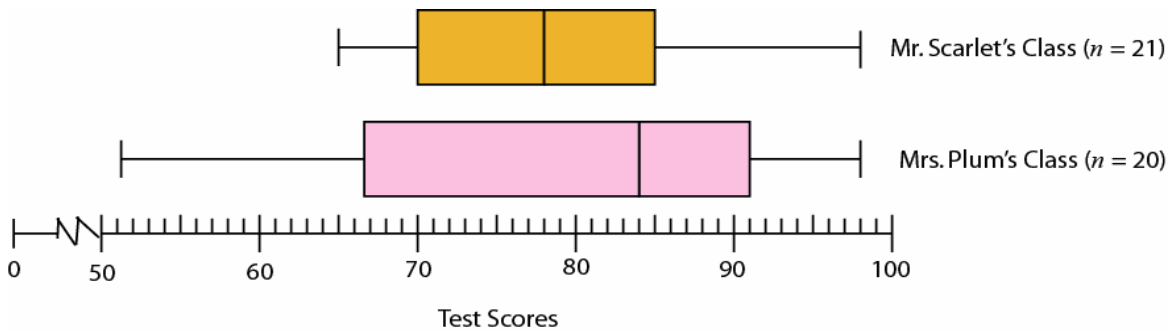
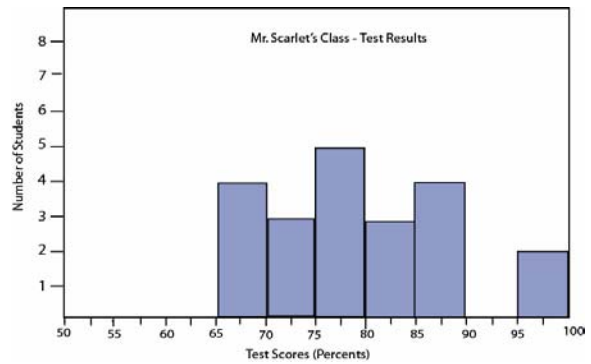
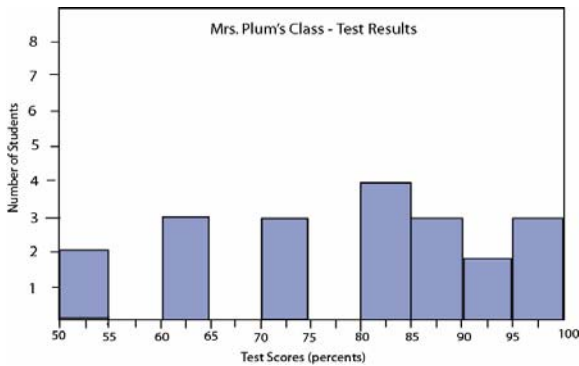
Using the data provided make a case for each of the following. Support each statement using the given data and distributions.

- 1) The students in both classes scored the same.
- 2) The students in Mrs. Plum’s class scored better than the students in Mr. Scarlet’s class.
- 3) The students in Mr. Scarlet’s class scored better than the students in Mrs. Plum’s class.

Data from Mathematics Test

Mean: Mrs. Plum’s class – about 78.8%; Mr. Scarlet – about 78.5%

Mode: Mrs. Plum’s class – 84%; Mr. Scarlet – 68, 70, and 76



(Definition *DSP* – 21 continued on following page)

Resource Material Prototype

Section 1: Data and Statistics

Sample Answer:

1) **The students in both classes scored the same:** The mean can be used to make the case that the classes scored the same since the mean scores for each class are within about 0.3%. Other measures including the mode (data given), range (different length whiskers), and median (on box-and-whisker) are all different. In addition, the shape of the histograms indicate considerable differences in the distributions of the scores across the range for both classes.

2) **The students in Mrs. Plum's class scored better than the students in Mr. Scarlet's class:**

The median score is higher in Mrs. Plum's class than in Mr. Scarlet's class, and more students scored above 90% in Mrs. Plum's class (3 more) than in Mr. Scarlet's class.

3) **The students in Mr. Scarlet's class scored better than the students in Mrs. Plum's class:**

While the median score was lower for Mr. Scarlet's class, the box-and-whisker plots and histograms can be used to make the case that the students in Mr. Scarlet's class scored better on the test.

- Five students (25%) scored below 65% in Mrs. Plum's class while no students in Mr. Scarlet's class scored below 65%.
- The range of scores was smaller for Mr. Scarlet's class (33%; from 65% to 98%) than Mrs. Plum's class (47%; from 51% to 98%).
- The histograms indicate that the scores in Mr. Scarlet's class are clumped closer together than the scores in Mrs. Plum's class. Therefore, there is less variation in the scores in Mr. Scarlet's class than Mrs. Plum's class.

A Note about Sample Answers...

This example illustrates that different conclusions can be drawn from the same data sets. But, more importantly, the sample answers show an understanding of the data and distributions, and exhibit the need to allow students to make different conclusions while defending their conclusions.

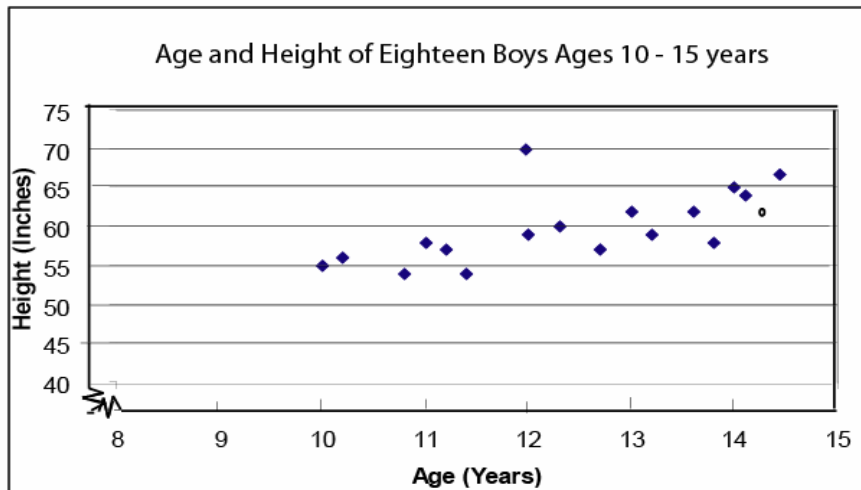
(Definition *DSP* – 21 continued on following page)

Resource Material Prototype

Section 1: Data and Statistics

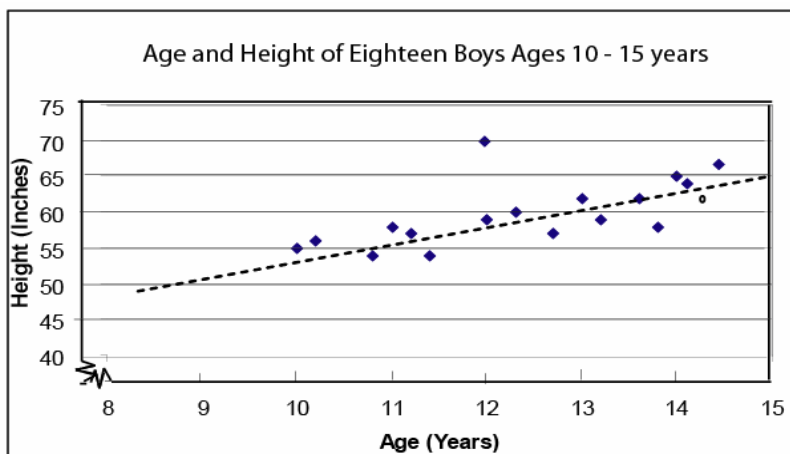
Example 21.8 – Makes a prediction:

Look at this scatter plot.



Using the data in this scatter plot *predict* the height of a typical 9 year old boy. Explain your answer.

Answer: about 50 inches tall; If an estimated line of best fit is placed on the graph to represent the trend between age and height and the trend line is extended to age 9, one could predict that a typical 9 year old might be approximately 50 inches tall.



(Definition *DSP* – 21 continued on following page)

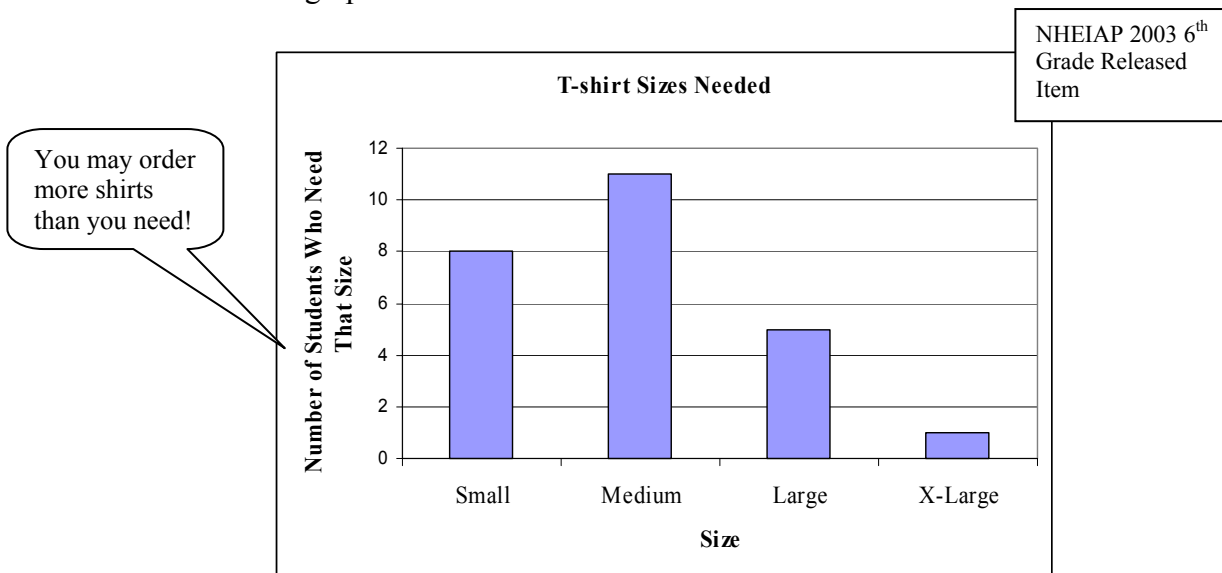
Resource Material Prototype

Section 1: Data and Statistics

Example 21.9 – Interprets a representation to solve a problem:

Use the graph below to answer this question.

Students in Ms. Gwin’s class want to purchase school T-shirts. The sizes they need are shown on the graph.



The prices for the T-shirts are

- \$6.00 for a package of 4, all the same size, or
- \$2.50 for each individual shirt.

a. How many packages and/or individual shirts should the students order to get both the **lowest price** possible and get **at least** one shirt for every student in the class? Be sure to list the sizes and the amounts ordered for each size. (Remember, you may order more shirts than you need.) Show your work and explain how you know your complete order has the lowest price.

b. What is the total cost for the complete order? Show your work or explain how you got your answer.

Answer:

- a) The students need to order at least 8 small, 11 medium, 5 large, and 1 X-large T-shirt.
The least expensive way to order the T-shirts is
2 packages of 4 small shirts = \$12.00
3 packages of 4 medium shirts = \$18.00 (2 packages and 3 individual shirts would cost \$19.50.)
1 package of 4 large shirts = \$6.00
1 individual large shirt = \$2.50
1 individual X-large shirt = \$2.50
b) TOTAL = \$41.00

(Definition *DSP* – 21 continued on following page)

Resource Material Prototype

Section 1: Data and Statistics

Example 21.10 – Interprets a representation to solve a problem involving mean:

Alika read a book on Monday, Tuesday, and Wednesday. He read an average of 10 pages per day. Indicate whether each of the following is possible or not possible by placing an X in the appropriate column.

1992 NAEP
released Item –
Grade 8

		Pages Read			
Possible	Not Possible		Monday	Tuesday	Wednesday
		A	4 pages	4 pages	2 pages
		B	9 pages	10 pages	11 pages
		C	5 pages	10 pages	15 pages
		D	10 pages	15 pages	20 pages

Answer:

		Pages Read			
Possible	Not Possible		Monday	Tuesday	Wednesday
	X	A	4 pages	4 pages	2 pages
X		B	9 pages	10 pages	11 pages
X		C	5 pages	10 pages	15 pages
	X	D	10 pages	15 pages	20 pages

- A) Not possible (average is $3\frac{1}{3}$ pages)
- B) Possible (average is 10 pages)
- C) Possible (average is 10 pages)
- D) Not possible (average is 15 pages)

Resource Material Prototype

Section 1: Data and Statistics

DSP – 22 Analyzes the impact of outliers on the mean, median and mode:

This GLE focuses on an understanding of the impact outliers have on measures of central tendency. This includes analyzing measures of central tendencies in sets of data to determine which set of data is most impacted by an outlier.

Example 22.1 – Analyzes impact of outliers on mean:

Mary and Mark were comparing their quiz scores.

Mary's quiz scores: 55%, 82%, 85%, 88%, 89%, 81%

Mark's quiz scores: 79%, 78%, 83%, 85%, 81%, 84%

If they each dropped their lowest score, whose average would make the most gain? Explain your answer without giving exact calculations.

Answer: Mary's average would make the greatest gain because all her scores are between 82% and 89% except for the 55% score; an outlier given all the other quiz scores. On the other hand all Mark's quiz scores range between 78% and 85%. Therefore, dropping the 78% will not have as great of an impact on Mark's score as the impact of dropping the 55% from Mary's scores. Mark's lowest score is typical of his other grades, not an outlier.

Example 22.2 – Analyzes impact of outliers on median:

Look at these data.

Mrs. Plum's Class	
5	1, 3
6	0, 1, 2
7	1, 2, 2
8	0, 4, 4, 4, 7, 8, 9
9	3, 4, 6, 7, 8

Mr. Scarlet's Class	
6	5, 8, 8, 8
7	0, 0, 0, 6, 6, 6, 8, 9
8	0, 1, 2, 5, 5, 8, 8
9	7, 8

Mrs. Plum dropped the two lowest scores on the test. Mr. Scarlet dropped the two highest scores on the test. Which class, Mrs. Plum's or Mr. Scarlet's, would see the greatest change in the median? Explain why.

Answer: The median would not change in Mrs. Plum's class since dropping the two lowest scores will shift the median up one position which does not change the median because of how the three scores of 84% are grouped around the original median. The median score in Mr. Scarlet's class would drop to 76% from 78% since dropping the two highest scores will shift the median down one position which would change the median because there are not multiple scores of 78% grouped around the median in the same way as Mrs. Plum's class.

Resource Material Prototype

Section 1: Data and Statistics

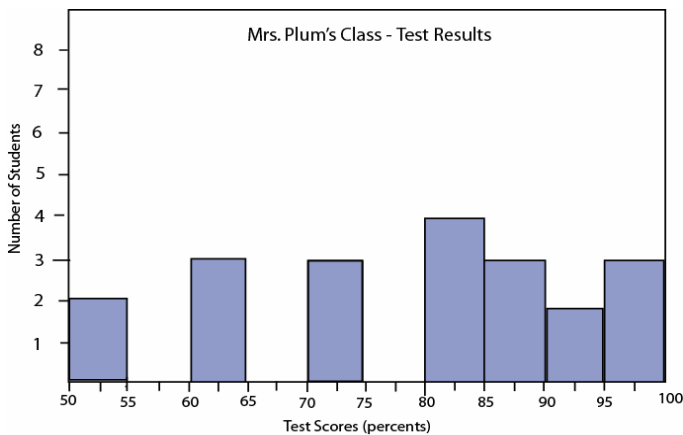
DSP – 23 Identifies or describes representations or elements of representations that best display a given set of data or situation: This stem focuses on student understanding of the best display to represent the situation. This includes analyses such as:

- the type of representation given a set of data (e.g., a bar graph for categorical data as opposed to a histogram for continuous data);
- the way in which the data is displayed (e.g., intervals in histogram are smaller to better represent spread and distribution);
- scales appropriate given the data (e.g., scales include the range of data and are spaced evenly); or
- explain why a graph is misleading.

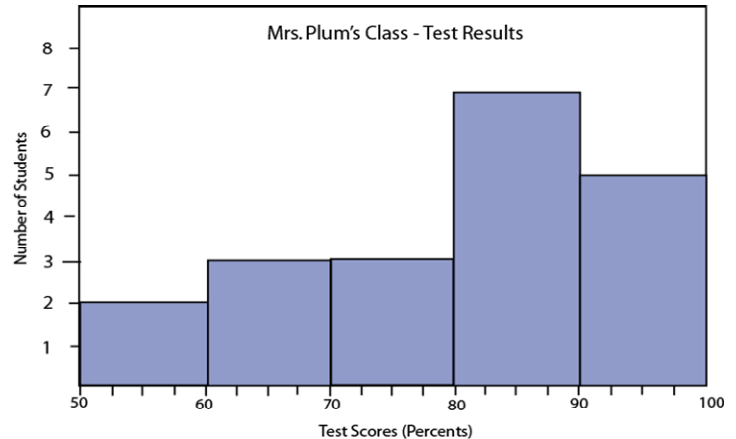
Example 23.1 – Element of a representation that best displays the data:

Both histograms below represent the same set of data from the test given by Mrs. Plum. The intervals designed in Display 1 better illustrate the distribution of the data than in Display 2.

What information can you get from the histogram in Display 1 that you can not get from the histogram in Display 2?



Display 1



Display 2

Answer: The shape of the data in each histogram is different because the data in Display 1 are broken into intervals of 5 percentage-points rather than intervals of 10 percentage-points as in Display 2. The finer grain in Display 1 allows one to see how the data are distributed within the 10 percentage-points. For example, in Display 2 one might conclude that the scores are distributed evenly across 50% – 60%, while in fact the two scores are between 50% and 55% as found in Display 1.

(Definition DSP – 23 continued on following page)

Resource Material Prototype

Section 1: Data and Statistics

Example 23.2 – Element of a representation that best displays the data (location of the dependent and independent variable reflect the situation, and the scales appropriately display the data to show the trend or pattern in the relationship):

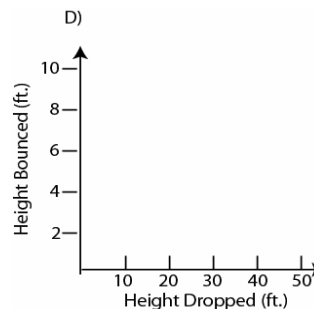
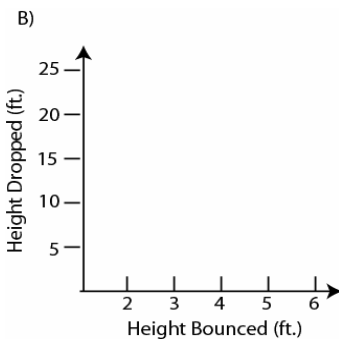
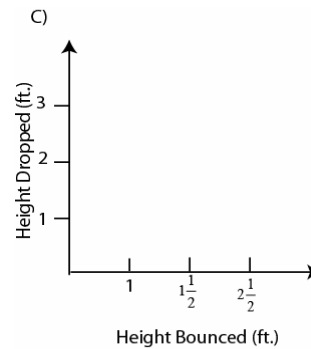
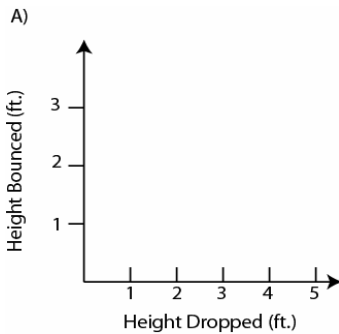
Use the chart below to answer this question.

NHEIAP Released 6th grade Item (2002)

Distance a Golf Ball Bounces

Height Dropped (ft.)	Height Bounced (ft.)
2	1
3	$1\frac{1}{2}$
5	$2\frac{1}{2}$

Brett did an experiment to observe how high a golf ball would bounce when dropped from different heights. The chart shows his results. Which of the following would be the best way for Brett to begin constructing a graph of the results? Explain why each representation would or would not be the best way to begin constructing a graph of the results.

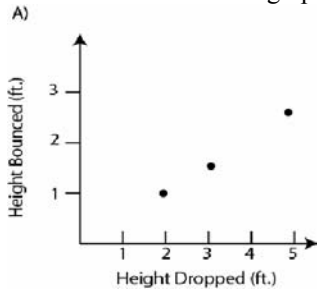


(Definition DSP – 23 continued on following page)

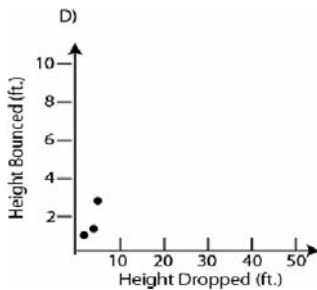
Resource Material Prototype

Section 1: Data and Statistics

Answer: A; Graph A represents the independent and dependent variable on their typical axes – independent on the horizontal axis and dependent on the vertical axis. Additionally, the following shows the data graphed using construction A.

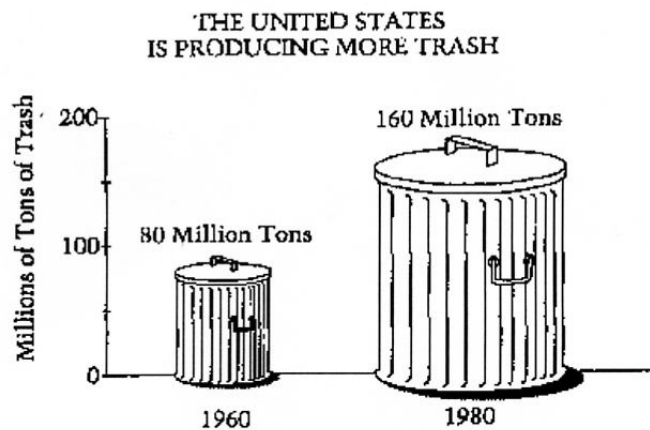


Graphs B, C, and D contain scaling flaws. Graph D will result in the data being clumped too closely as shown below.



The horizontal axis on Graph C has the values, 1, $1\frac{1}{2}$, $2\frac{1}{2}$, incorrectly evenly spaced while the scale for Graph B will result in clumping data as in Graph D.

Example 23.3 – Element of a representation that is misleading:



NAEP 1992
Released Item –
Grade 8

The pictograph shown above is misleading. Explain why.

Answer: Since both the diameter and height was doubled, it seems that the change is greater than it is.

(Definition DSP – 23 continued on following page)

Resource Material Prototype

Section 1: Data and Statistics

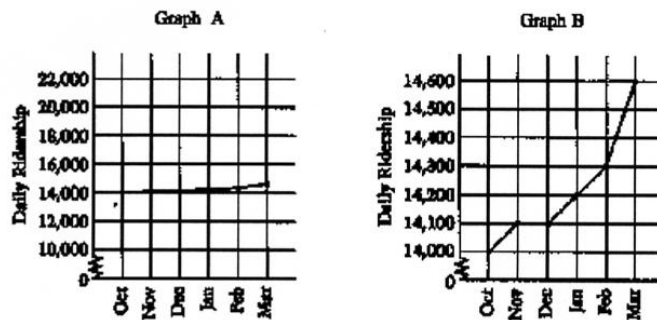
Example 23.4 – Element of a representation that is misleading:

This question requires you to show your work and explain your reasoning. You may use drawings, words, and numbers in your explanation. Your answer should be clear enough so that another person could read it and understand your thinking. It is important you show all of your work.

METRO RAIL COMPANY	
Month	Daily Ridership
October	14,000
November	14,100
December	14,100
January	14,200
February	14,300
March	14,600

NAEP 1996 Released
Item – Grade 8

The data in the table above have been correctly represented by both graphs shown below.



- 1) Which graph would be best to help convince others that the Metro Rail Company made a lot more money from ticket sales in March than in October?
- 2) Under what circumstances might Graph A be misleading? Graph B? Explain your answer.

Answer:

Question 1: Graph B; Graph B appears to show a larger increase from October to March than Graph A.

Question 2: If a 4% increase is a significant increase for the Metro Rail Company, Graph A seems to misrepresent this increase. Graph B seems to better convey this increase. The opposite would be true if a 4% increase is not considered significant.

Resource Material Prototype

Section 1: Data and Statistics

DSP – 24 Evaluates sample from which the statistics were developed

(bias): This aspect of GLE M(DSP)–X–2 focuses on student understanding of bias in a sample.

NAEP 2003
Released Item –
Grade 8

Example 24.1:

A survey is to be taken in a city to determine the most popular sport. Would sampling opinions at a baseball game be a good way to collect these data? Explain your answer.

Answer: No; opinions may be biased in favor of those who like baseball.

NAEP 1996
Released Item –
Grade 8

Example 24.2:

A poll is being taken at Baker Junior High School to determine whether to change the school mascot. Which of the following would be the best place to find a sample of students to interview that would be most representative of the entire student body?

- A) An algebra class
- B) The cafeteria
- C) The guidance office
- D) A French class
- E) The faculty room

Answer: B; The cafeteria because it would likely have the best distribution of students throughout the school, while the other groups of students might reflect a grade in the school, or specific interests students have.

Resource Material Prototype

Section 2: Counting Principles and Probability

NECAP: M(DSP) – X – 4, M(DSP) – X – 5

Vermont: MX: 26, MX: 27

Definition	Page Number	Definition Number
Combination	43	<i>DSP – 26</i>
Event	47	<i>DSP – 31</i>
Experimental probability	50	<i>DSP – 33</i>
Fundamental counting principle	46	<i>DSP – 29</i>
Permutation	44	<i>DSP – 27</i>
Sample space	47	<i>DSP – 30</i>
Solves problems using a variety of counting strategies	41	<i>DSP – 25</i>
Theoretical probability	48	<i>DSP – 32</i>
Tree diagram	45	<i>DSP – 28</i>

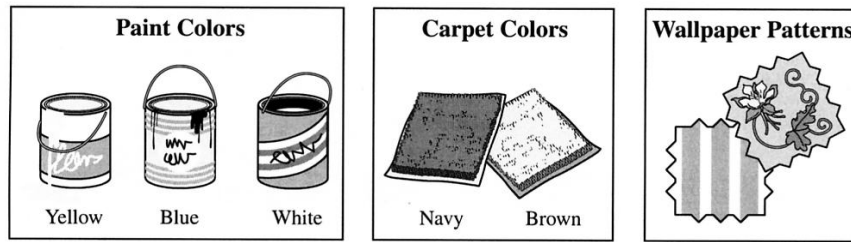
Resource Material Prototype

Section 2: Counting Principles and Probability

DSP – 25 Solves problems using a variety of counting strategies: To apply a variety of counting strategies such as organized lists, tables, and tree diagrams to systematically solve counting problems.

Example 25.1 – Using a table to solve a counting problem:

Darnae is planning to decorate her room. She will choose 1 color of paint, 1 color of carpet, and 1 pattern for wallpaper. The colors and patterns she will choose from are shown below.



MCAS – Grade 4
Released Item 2002

How many ways are there to choose 1 paint color, 1 carpet color, and 1 wallpaper pattern?

- A. 7
- B. 8
- C. 10
- D. 12

Answer: D; See the table below for the 12 different ways to choose 1 paint color, 1 carpet color, and 1 wallpaper pattern.

Paint Colors			Carpet Colors		Wallpaper Patterns	
Yellow	Blue	White	Navy	Brown	Striped	Flowers
x			x		x	
x			x			x
x				x	x	
x				x		x
	x		x		x	
	x		x			x
	x			x	x	
	x			x		x
		x	x		x	
		x	x			x
		x		x	x	
		x		x		x

(Definition DSP – 25 continued on following page)

Resource Material Prototype

Section 2: Counting Principles and Probability

Example 25.2 – Using a diagram to solve a counting problem:

Martin wants to buy some animal crackers. The sign on the machine says:

EACH ITEM – 50¢.
USE NICKELS, DIMES, AND QUARTERS.
USE EXACT CHANGE ONLY.

Martin decides that one way to pay for the crackers is to use 4 dimes and 2 nickels.

a. What other ways can Martin pay for the animal crackers? Show as many ways as possible that he can use nickels, dimes, and quarters to buy the animal crackers.

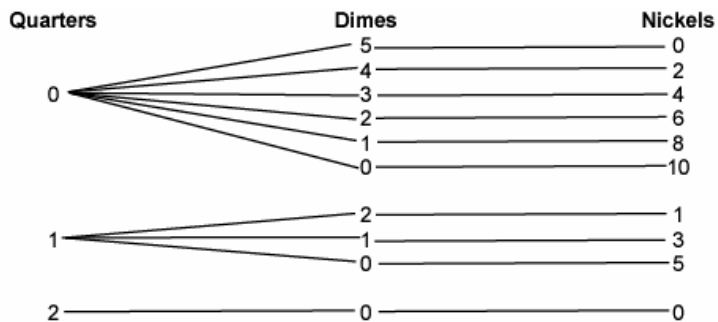
b. Draw and label the fewest coins Martin can use.

NH 2003 Released Item – Grade 3

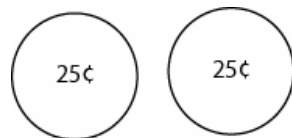
c. Draw and label the most coins Martin can use.

Answer:

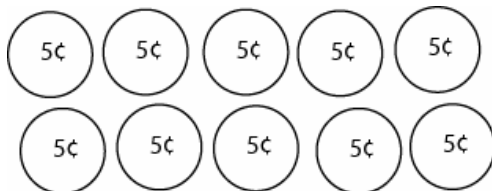
a)



b)



c)



(Definition DSP – 25 continued on following page)

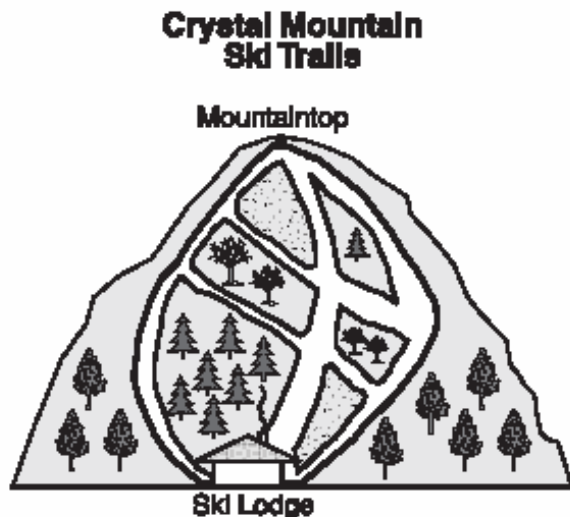
Resource Material Prototype

Section 2: Counting Principles and Probability

Example 25.3 – Solving a problem involving different paths:

Use the diagram to answer this question.

NH Released Item 2003 – Grade 6



Seth is at the top of Crystal Mountain. How many different routes can he take to get to the ski lodge if he only goes downhill?

- A. 3
- B. 5
- C. 10
- D. 12

Answer: C

DSP – 26 Combination: The number of ways that a set of items can be selected from a set of distinguishable items when *order does not matter*.

Example 26.1 – Choosing three objects from a set of five when order does not matter:

Mr. Relihan’s class is conducting a mock trial. Mr. Relihan will select a jury of 3 students from a group of 5 students – Jan, Linda, Tim, Ken, and Debbie. In how many ways can Mr. Relihan select the jury? Show your work or explain how you know.

Answer: 10 ways; Mr. Relihan can pick the first person in any of 5 ways, then the second person in any of 4 ways, and the third person in any of 3 ways. Since the order of selection does not matter and any three names can be arranged in 6 different ways (e.g., if Jan, Linda, and Ken are chosen, their names can be arranged in six different ways, but each of these ways still forms the same three-person jury), the total number of ways to choose the jury is $\frac{(5)(4)(3)}{6} = 10$.

Resource Material Prototype

Section 2: Counting Principles and Probability

DSP – 27 Permutation: The number of ways that a set of items can be selected from a set of distinguishable items when *order matters*.

Example 27.1 – Number of permutations of 3 objects taking 3 at a time:

The three digits in an area code are 1, 5, and 8. What are all the possible arrangements of these digits that could be the area code?

Answer: There are six possibilities for arranging the digits 1, 5, and 8: 158, 185, 518, 581, 815, and 851.

Example 27.2 – Number of permutations of 5 objects taking 2 at a time:

A club has 5 members: Alisa, Rich, Marge, Dick, and Quinn. The club has a president and a treasurer. How many different ways can the president and treasurer be selected from the 5 club members? Show your work or explain how you know.

Answer: There are twenty different ways that the president and treasurer can be selected from the 5 club members. The table below accounts for each selection of president or treasurer (e.g., AR means Alisa is president and Rich is the treasurer, while RA means Rich is president and Alisa is treasurer).

Treasurer

	A	R	M	D	Q
A		AR	AM	AD	AQ
R	RA		RM	RD	RQ
M	MA	MR		MD	MQ
D	DA	DR	DM		DQ
Q	QA	QR	QM	QD	

Key

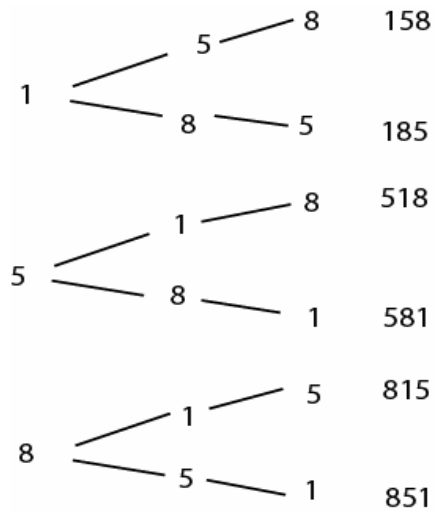
A represents Alisa; R represents Rich; M represents Marge; D represents Dick; and Q represents Quinn.

Resource Material Prototype

Section 2: Counting Principles and Probability

DSP – 28 Tree diagram: A tree diagram is a strategy that can be used for identifying combinations, permutations, or as a general counting technique for solving problems and is often used in conjunction with the Fundamental Counting Principle (See *DSP – 28.*).

Example 28.1: (See Example 27.1.) The three digits in an area code are 1, 5, and 8. What are all the possible arrangements of these digits that could be the area code?



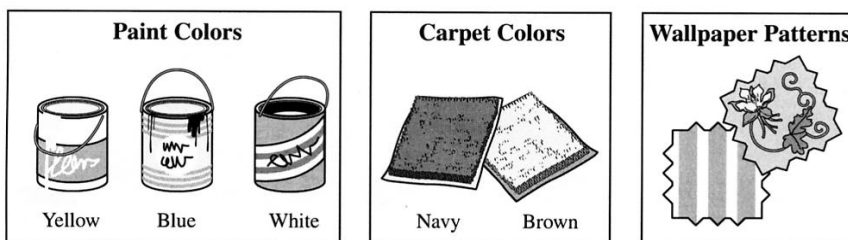
Resource Material Prototype

Section 2: Counting Principles and Probability

DSP – 29 Fundamental Counting Principle: A rule that states if there are x_1 ways to choose a first item, x_2 ways to choose a second item, x_3 ways to choose a third item, and so on, the total number of ways to choose all the items is determined by multiplying the number of ways to choose each item by each other.

Example 29.1 – Application of the Fundamental Counting Principle:

Darnae is planning to decorate her room. She will choose 1 color of paint, 1 color of carpet, and 1 pattern for wallpaper. The colors and patterns she will choose from are shown below.



MCAS 2002 Grade
4 Released Item

How many ways are there to choose 1 paint color, 1 carpet color, and 1 wallpaper pattern?

- A. 7
- B. 8
- C. 10
- D. 12

Answer: D; There are 3 paint colors (x_1), 2 carpet colors (x_2), and 2 wallpaper designs (x_3). Therefore, multiply $3 \cdot 2 \cdot 2$ to determine that there are 12 different ways to choose 1 paint color, 1 carpet color, and 1 wallpaper pattern.

Resource Material Prototype

Section 2: Counting Principles and Probability

DSP – 30 Sample space: Sample space is the set of all the possible outcomes for an experiment.

Example 30.1 – Sample space for rolling a single six-sided number cube:



$\{1, 2, 3, 4, 5, 6\}$

Example 30.2 – Sample space for rolling a pair of six-sided number cubes:



Values on First
Number Cube

Values on Second Number Cube

	1	2	3	4	5	6
1	1,1	1,2	1,3	1,4	1,5	1,6
2	2,1	2,2	2,3	2,4	2,5	2,6
3	3,1	3,2	3,3	3,4	3,5	3,6
4	4,1	4,2	4,3	4,4	4,5	4,6
5	5,1	5,2	5,3	5,4	5,5	5,6
6	6,1	6,2	6,3	6,4	6,5	6,6

The sample space for rolling a pair of six-sided number cubes contains the set of 36 pairs of numbers found in this table.

Example 30.3 – Sample space for tossing a coin:

$\{\text{Heads, Tails}\}$



DSP – 31 Event: A subset of a sample space (i.e., a set of possible outcomes).

Example 31.1 – Tossing a six-sided number cube:

The following are examples of different events (subsets of the sample space) when you toss a six-sided number cube once.

- A) The event that we roll an even number is the subset $\{2, 4, 6\}$ of the sample space $\{1, 2, 3, 4, 5, 6\}$.
- B) The event that we roll the number 5 is the subset $\{5\}$ of the sample space $\{1, 2, 3, 4, 5, 6\}$.
- C) The event that we roll an odd number is the subset $\{1, 3, 5\}$ of the sample space $\{1, 2, 3, 4, 5, 6\}$.

Resource Material Prototype

Section 2: Counting Principles and Probability

DSP – 32 Theoretical probability: Theoretical probability is the likelihood of an event occurring based upon the number of ways an event can occur and the total number of possible outcomes in a sample space of equally likely outcomes. The probability of an event is a number between 0 and 1. The closer the probability is to 1, the more likely the event will occur. (See **DSP – 31** for a definition of Event.)

The probability of Event A is the number of ways that A can occur divided by the total number of possible outcomes (in a sample space of equally likely outcomes).

$$P(A) = \frac{\text{The number of ways Event A can occur}}{\text{The total number of possible outcomes}}$$

Example 32.1 – Outcomes that are equally likely to occur:

A single six-sided number cube is rolled. What is the probability of rolling a 2?

Answer: Since the sample space for rolling a number cube is $\{1, 2, 3, 4, 5, 6\}$, there are six possible equally likely outcomes. Therefore, the probability of rolling a 2 is $\frac{1}{6}$.

The probability of rolling a 2.

$$P(2) = \frac{\text{Number of ways to roll a 2}}{\text{Total number of outcomes}} = \frac{1}{6}$$

Example 32.2 – Outcomes that are equally likely to occur:

A six-sided number cube is rolled. What is the probability of rolling a multiple of 3?

Answer: Since the sample space for rolling a six-sided number cube is $\{1, 2, 3, 4, 5, 6\}$, there are six equally likely outcomes. Since rolling a multiple of 3 can occur in two ways, rolling a 3 or rolling a 6, the probability of rolling a multiple of 3 is $\frac{2}{6}$ or $\frac{1}{3}$.

The probability of rolling a multiple of 3, $\{3, 6\}$.

$$P(\text{multiple of 3}) = \frac{\text{Number of ways to roll a multiple of 3}}{\text{Total number of outcomes}} = \frac{2}{6} = \frac{1}{3}$$

If an outcome of this experiment is a 3 or a 6, then we say the event, rolling a multiple of 3, or $\{3, 6\}$, has occurred.

(Definition **DSP – 32** continued on following page)

Resource Material Prototype

Section 2: Counting Principles and Probability

Example 32.3 – Outcomes that are NOT equally likely to occur:

A bag contains 4 red, 3 blue, 8 white, and 5 yellow marbles. What is the probability of randomly selecting 1 red marble from the bag?

Answer: If we consider the sample space to be {red, blue, white, yellow} then the outcomes are not equally likely since the numbers of each colored marble are different. In this case definition *DSP – 32* does not directly apply. However, we may consider the sample space to contain 20 equally likely outcomes by naming each marble (e.g., {red 1, red 2, red 3, red 4, blue 1, and so on}). Then *DSP – 32* directly applies, and the probability can be calculated as follows.

$$P(\text{red}) = \frac{\text{Number of ways to select 1 red marble}}{\text{Total number of marbles}} = \frac{4}{20} = \frac{1}{5} = 0.20 = 20\%$$

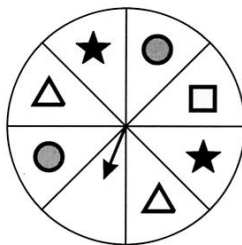
The probability the event selecting 1 red marble

Probabilities can be expressed as fractions, decimals, percents, ratio, or, in this case, in the form of one out of five.

Example 32.4 – Establishing equally likely outcomes:

CURRICULUM GOALS 5, 6, 7, & 8

Use the picture of the spinner below to answer question 8.



8. In order to have an **EQUAL** chance of landing on each shape, which shape should go in the blank space?

- A.
- B.
- C.
- * D.

NHEIAP Released
Item 2001–02

Resource Material Prototype

Section 2: Counting Principles and Probability

DSP – 33 Experimental probability: The probability of an event occurring based upon repeated testing. That is, if an experiment is repeated k times and an event is observed n times, then the experimental probability of the event occurring is $\frac{k}{n}$.

Example 33.1 – Tossing a coin:

If a coin is tossed 2000 times, n , and heads is observed 989 times, k , then the experimental probability is $\frac{k}{n} = \frac{989}{2000} = 0.4945$.

Resource Material Prototype

References:

Agrawal, Piyush C. (1993). *Mathematics Applications and Connections*. Lake Forest, IL: Macmillan/McGraw-Hill.

Billstein, Rick, and Williamson, Jim. (1999). *MATH Thematics*. Evanston, Illinois: McDougal Littell.

Borowski, E.J., and Borwein, J.M. (1991). *The HarperCollins Dictionary of Mathematics*. New York, NY: HarperCollins Publishers.

Cavanagh, Mary C. (2000). *Math To Know*. Wilmington, MA: Great Source Education Group.

Freedman, D., Pisani, R., and Purves, R. (1997) *Statistics, 3rd. Edition*. W.W. Norton and Company.

Data and Story Library. Ice Cream Data. (1953).
<http://lib.stat.cmu.edu/DASL/Stories/Ageandheight.html>.

Lamon, S.J. *Teaching Fractions and Ratios for Understanding*. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Larson, Roland E., Kanold, Timothy D., and Stiff, Lee. (1995). *Algebra 1: An Integrated Approach*. Lexington, MA: D.C. Heath and Company.

Licker, Mark D. (2003). *Dictionary of Mathematics*. New York, NY: McGraw Hill.

Lilly, Marsha. (1999). *Math at Hand*. Wilmington, MA: Great Source Education Group.

Marsden, Ted. (2004). (Conversations), Norwich University, Northfield, Vermont.

Mathworld, Wolfram, <http://mathworld.wolfram.com/>.

Massachusetts Department of Education, MCAS, 2003 Released Item

National Center for Health Statistics Growth Chart.
<http://www.cdc.gov/nchs/data/nhanes/growthcharts/set2/chart%2007.pdf>.

National Assessment of Educational Progress released items.
<http://nces.ed.gov/nationsreportcard/itmrls/>.

Nelson, David. (2003). *The Penguin Dictionary of Mathematics*. London, England: Penguin Books.

Resource Material Prototype

Normal Daily Maximum Temperatures.

<http://www.met.utah.edu/jhorel/html/wx/climate/maxtemp.html>.

Rosenfeld, R., discussions and access to an unpublished work, Vermont Mathematics Initiative.

Shodor Interactivate, Dictionary, <http://www.shodor.org/interactivate/index.html>.

Spiegel, Murray R. 1992. *Schaum's Outline Series, Theory and Probability and Statistics*. McGraw-Hill, Inc.

University of Chicago School Mathematics Project. (1993). *Algebra*. Glenview, IL: ScottForesman.

University of Chicago School Mathematics Project. (2004). *Everyday Mathematics; Teacher's Reference Manual*. Columbus, OH: McGraw Hill SRA.

Usiskin, Zalman. (1992). *Transition Mathematics*. Glenview, IL: Scott, Foresman.

Van De Walle, John A. (2001). *Elementary and Middle School Mathematics: Teaching Developmentally*. New York, NY: Addison Wesley Longman, Inc.