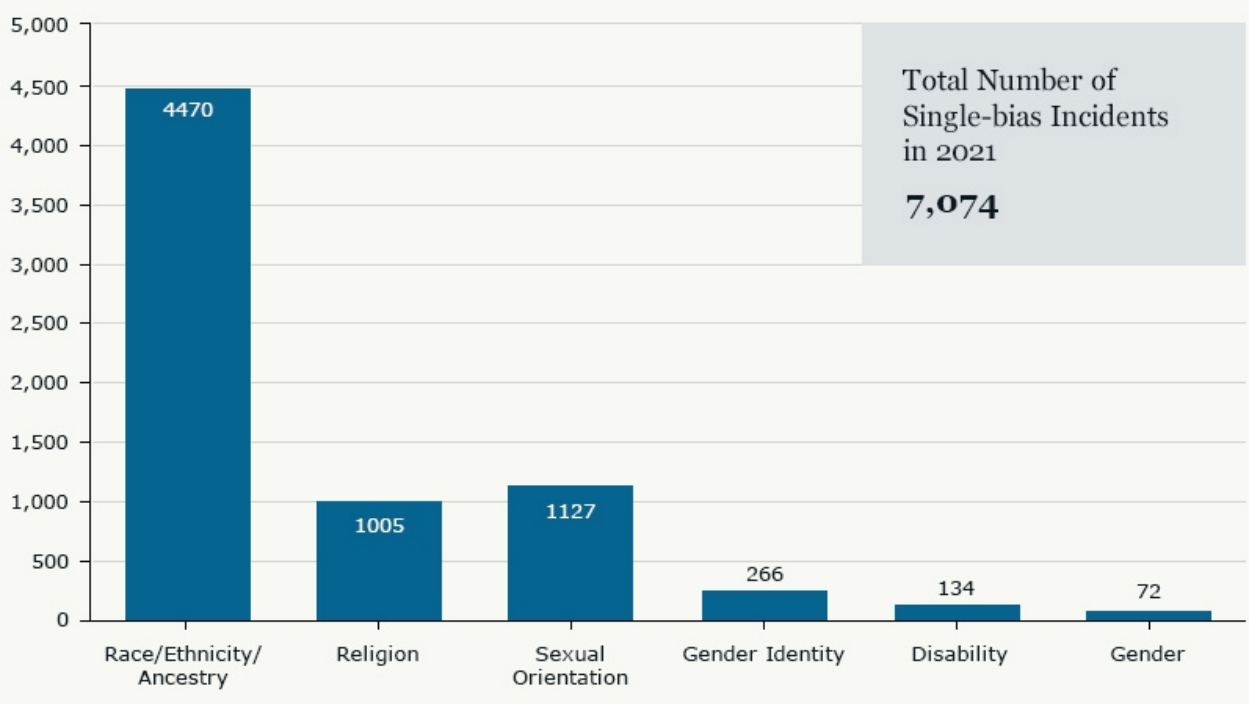
Chapter 6 BASIC STATISTICS



https://www.justice.gov/hatecrimes/hate-crime-statistics

In this chapter…

Introduction

Although statistics is not typically included in texts based on algebra, we include it in this text as a supplement for future elementary school teachers since they will teach statistical reasoning in their classrooms. This chapter may also be used as a module supplement in other courses and/or curricula designed for future elementary school teachers.

Data analysis is addressed throughout elementary school, beginning with grade 1. Below are the related standards from the Massachusetts Mathematics Curriculum Framework for K - 6.

Grade 1 (1.MD.C) Represent and interpret data.

4. Organize, represent, and interpret data with up to three categories; ask and answer questions about the total number of data points, how many in each category, and how many more or less are in one category than in another.

Grade 2 (2.MD.D) Represent and interpret data.

9. Generate measurement data by measuring lengths of several objects to the nearest whole unit, or by making repeated measurements of the same object. Organize and record the data on a line plot (dot plot) where the horizontal scale is marked off in whole-number units.

10. Draw a picture graph and a bar graph (with single-unit scale) to represent a data set with up to four categories. Solve simple put-together, take-apart, and compare problems using information presented in a bar graph.

Grade 3 (3.MD.B) Represent and interpret data

3. Draw a scaled picture graph and a scaled bar graph to represent a data set with several categories. Solve one- and two-step “how many more” and “how many less” problems using information presented in scaled bar graphs. For example, draw a bar graph in which each square in the bar graph might represent five pets.

4. Generate measurement data by measuring lengths of objects using rulers marked with halves and fourths of an inch. Record and show the data by making a line plot (dot plot), where the horizontal scale is marked off in appropriate units—whole numbers, halves, or fourths.

Grade 4 (4.MD.B) Represent and interpret data

4. Make a line plot (dot plot) representation to display a data set of measurements in fractions of a unit (1⁄2, 1⁄4, 1⁄8). Solve problems involving addition and subtraction of fractions by using information presented in line plots (dot plots). For example, from a line plot (dot plot) find and interpret the difference in length between the longest and shortest specimens in an insect collection.

Grade 5 (5.MD.B) B. Represent and interpret data.

2. Make a line plot (dot plot) to display a data set of measurements in fractions of a unit. Use operations on fractions for this grade to solve problems involving information presented in line plot (dot plot). For example, given different measurements of liquid in identical beakers, find the amount of liquid each beaker would contain if the total amount in all the beakers were redistributed equally.

These standards address collecting, displaying and doing basic analysis of data. This knowledge prepares them for the study of measures of center and data spread in middle school. See statistics standards for grade 6 below.

Grade 6 (6.SP) Statistics and Probability

A. Develop understanding of statistical variability.

1. Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. For example, “How old am I?” is not a statistical question, but “How old are the students in my school?” is a statistical question because one anticipates variability in students’ ages.
2. Understand that a set of data collected to answer a statistical question has a distribution, which can be described by its center (median, mean, and/or mode), spread (range, interquartile range), and overall shape.
3. Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

B. Summarize and describe distributions.

4. Display numerical data in plots on a number line, including dot plots, histograms, and box plots. Read and interpret circle graphs.
5. Summarize numerical data sets in relation to their context, such as by:
    a. Reporting the number of observations.
    b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.
    c. Giving quantitative measures of center (median, and/or mean) and variability (range and/or interquartile range), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
    d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

In this chapter we introduce basic statistics as one might see it in an elementary classroom and beyond.

6.1 Data Samples

The idea of data sampling is surprisingly complex but groundwork for it can be laid in elementary school. The NCTM Standards advocates that students in **grades 3-5** should…

> “…*begin to understand that many data sets are samples of larger populations. They can look at several samples drawn from the same population, such as different classrooms in their school, or compare statistics about their own sample to known parameters for a larger population… they can think about the issues that affect the representativeness of a sample – how well it represents the population from which it is drawn – and begin to notice how samples from the population can vary* (p. 181).”

We will address some of the terminology in this quote later when we get into sampling but first we briefly address data types.

Broadly speaking, there are two types of data:

numerical and *categorical*

Numerical data (also known as *quantitative data*): Data as a measurement, such as a person's height or weight. Such data may also be a count, such as the number of cookies a person ate or hours spent on video games (line plot illustration below).

Image above from https://yorkfourthgrade.weebly.com/1-data.html

Categorical data (also known as *qualitative data*): Data as a characteristic such as a person's hair color, hometown, types of movies they like or sexual orientation (bar graph illustration below).

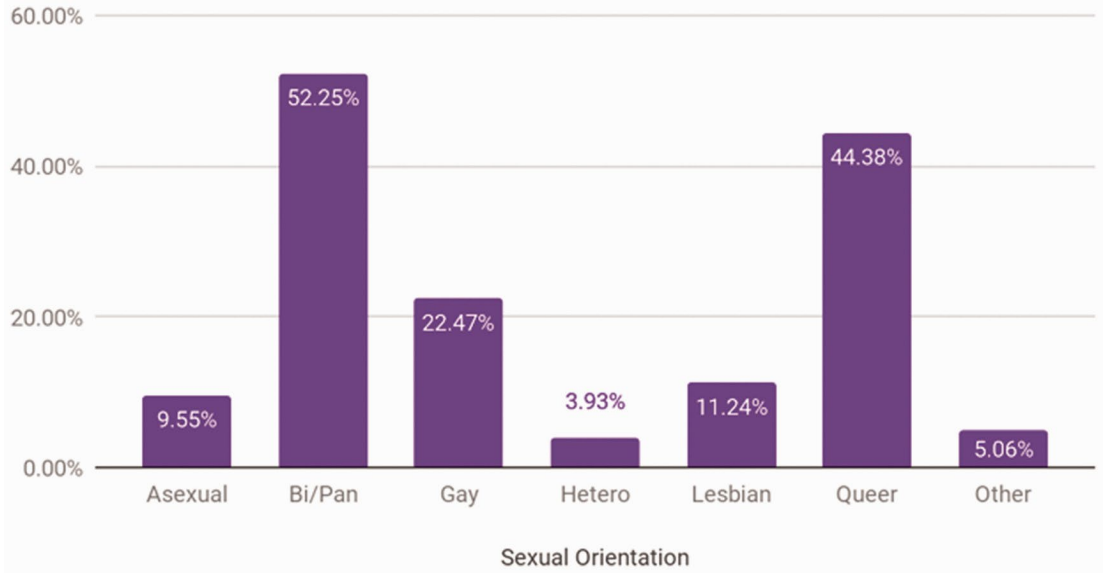


Image above from article at https://journals.sagepub.com/doi/10.1177/2053951720933286

No matter what the data type, it is always important to consider how the data were selected. We illustrate this with the following activity.

Activity: Consider the population of 100 rectangles on the next page. Notice that each one has an area (a size). For example, rectangle number 74 has size 6 square units because it is composed of 6 small squares.

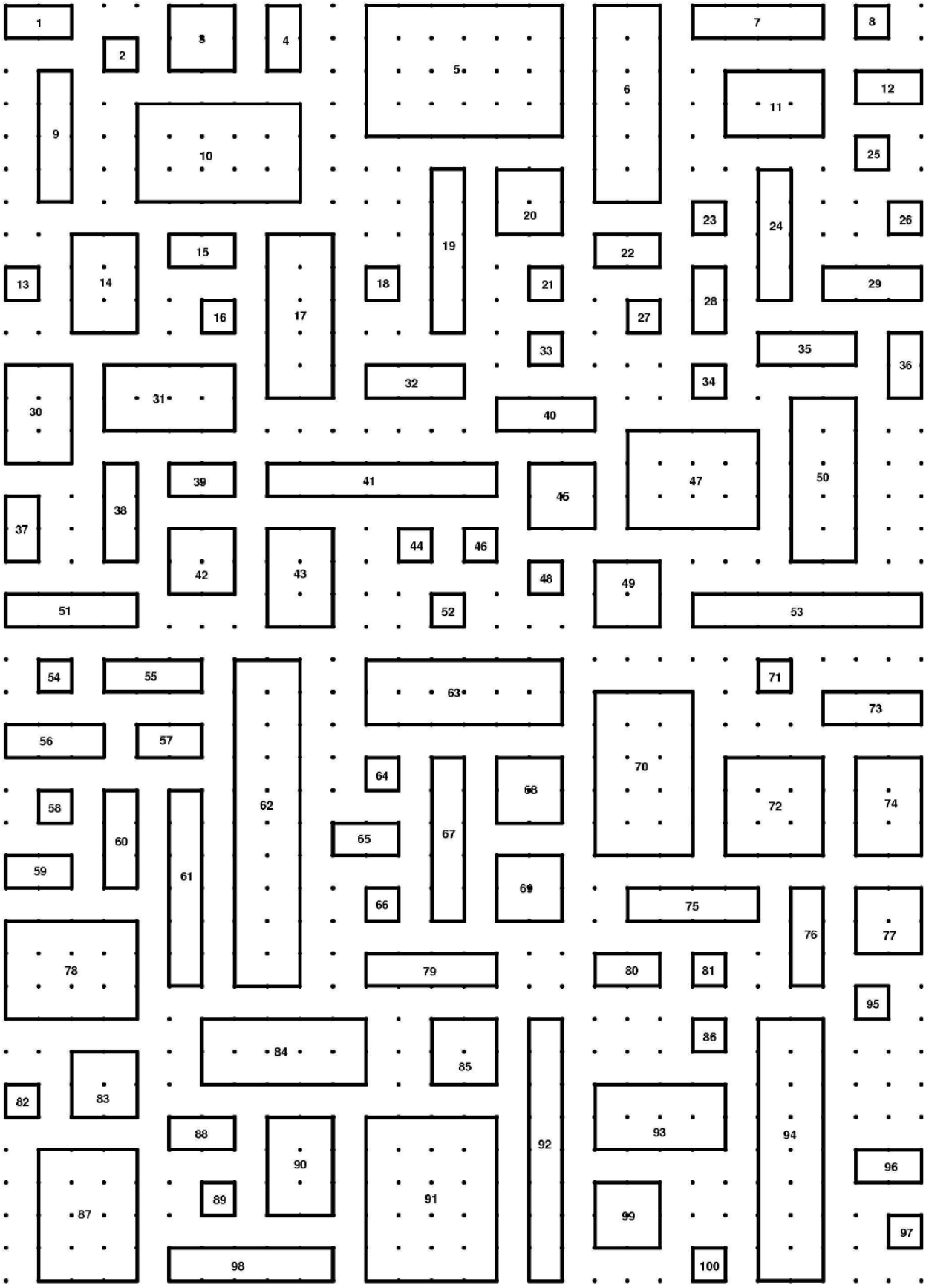
1) Each student carefully select a sample of ten rectangles that they think will have an average size close to the average size of all 100 rectangles on the page. Find the average size of your self-selected ten. Record the areas of your chosen rectangles and the average area below. Then, as a class, record averages on the board as follows: make a number line on the board that shows each distinct average in numerical order. (Each student) Put an x above your average number. We call this a line plot of the data.

2) Each student selects 10 rectangles at *random*. This will be done as follows: Your instructor will have a bowl with each of the 100 rectangle areas written on its own piece of paper and each student will have a turn to pick 10 pieces of paper at random. After each student picks and records their numbers, they will put them back into the bowl and the teacher will mix up the papers before the next student picks their 10. Write your 10 areas below and calculate the average area of these. Then, as a class, make a line plot of everyone's averages on the board.

3) Estimate the average size of all 100 rectangles, based on the line plots on the board. Write your answer below and explain why you chose the number you did.

4) As a class, calculate the actual average area of all 100 rectangles by splitting up the numbers to add into groups (e.g. 10 groups of 10 rectangles for 10 students to add up areas) and then adding up all of your subtotals and finally dividing by 100. Record the result below. Discuss how close or far this number is from your estimates in #1, 2 and 3.

A Population of Rectangles



This rectangle size problem contains a big lesson in statistics: Humans are not always good at selecting samples. If you want unbiased information, it is best to choose samples at random. In order to talk more about this, we need to discuss the language behind the sampling of data.

A **population** is the entire group that the collected data are intended to describe.

A **sample** is a subset of the population from which data are collected.

A **parameter** is a value calculated using *all* the data from the population. Parameters represent a property of the *population* (e.g. average height).

A **statistic** is a value calculated using data from a sample. Statistics represent a property of the *sample* (e.g. average height).

The **sampling rate** is calculated as follows: [(# in the sample) ÷ (# in the population)] This number is often multiplied by 100 to get a percent.

Example 6.1.1: In our rectangle problem, the population was the set of all 100 rectangles and we took samples of size 10.

a) What is the sampling rate?

b) What is the parameter average?

c) What is the statistic average from your random sample?

Example 6.1.2

According to Data USA, in 2021, of the 13,348 full time students enrolled at the Fitchburg State University, there were 1280 African Americans, 1386 Hispanics or Latinos and 406 Asians. If you were to take a random sample of 500 students from that population, approximately how many of each ethnic group would you expect to find in your sample?
Hint: First, find what percent of the population each ethnic group is.

African Americans:

Hispanics:

Latinos:

Asians:

A sample is **biased** if members of the population do not have equal likelihood of being in the sample.

For example if we wanted to gather data on weights of students at a particular school, the following would be considered biased samples.

Students in a particular class
Students in a particular grade

Discuss why data from these samples may be biased.

Can you think of other samples that might be biased?

Types of Bias

Sampling bias: when the sample is not representative of the population
Voluntary response bias: the sampling bias that often occurs when the sample is volunteers
Self-interest bias: bias that can occur when the researchers and/or participants have an interest in the outcome
Response bias: when the responder gives inaccurate responses for any reason
Perceived lack of anonymity: when the responder fears giving an honest answer might negatively affect them
Loaded question bias: when the question wording influences the responses
Non-response bias: when people refusing to participate in the study can influence the validity of the outcome

Example 6.1.3

In each situation, identify a potential source of bias. Discuss.

a) A survey asks how many sexual partners a person has had in the last year.

b) A radio station asks readers to phone in their choice in a daily poll.

c) A substitute teacher wants to know how students in the class did on their last test. The teacher asks the 10 students sitting in the front row to state their latest test score.

d) High school students are asked to report if they have consumed alcohol in the last two weeks.

e) The Beef Council releases a study stating that consuming red meat poses little cardiovascular risk.

f) A poll asks, “Do you prefer to shop at Walmart or spend more to shop at Target?”

If we picked a sample at random from all the students at a school then this would be an unbiased sample of students at that particular school. This is what we refer to as **simple random sampling** - like picking names out of a hat.

There is another type of random sampling called **cluster (random) sampling** where we break up the population into groups/clusters and randomly select a certain number from each group. For example if we randomly choose 20 students from each grade at the school to collect weight data.

Yet another type of sampling is called **systematic sampling** in which researchers choose members of the population at regular intervals. For example, if we pick every 10th person in a population of 100 then this would be considered a systematic random sampling. Sources vary on whether or not systematic sampling is considered random. What do you think? **Discuss**

The point of a statistical study is to use a sample to learn about a population by computing statistics which approximate the population’s true parameters. However, even a good representative sample will only approximate the population’s actual parameters.

Sampling error is based on the difference between the parameter and the statistic. We will not formally define it in this class, but understanding the idea provides a good start for elementary school students to appreciate its importance. In practice, we often do not know what the sampling error is because we don’t have the means to gather information about the entire population in order to find parameters (hence the point of sampling).

However, it makes sense that in order **to reduce sampling error we should make our sample random and big enough.**

So, what is “big enough”?
It seems like you would need some significant sampling rate (like, maybe 25% of the population), but it turns out that good surveys can be done with quite small samples. For example, national polling to predict the outcome of the presidential election is often done with only a few thousand people. That’s a few *thousand* out of more than 100 million likely voters (0.003% of the population). The sampling rate can be so small because people are not all that diverse from a sampling perspective.

Think of it this way: suppose you are going to sample a big pot of soup to see how it tastes. As long as that soup is stirred up really well, and the chunks of meat and vegetables in there aren’t too big or too different, all you need is a small bite to know the flavor.

Example 6.1.4

In order to study how satisfied the 41,732 citizens of Fitchburg, MA are with snow removal from city streets, I conducted the following study. I stood in front of Market Basket on a Monday morning in February from 9:00am until noon and I asked every third person who walked in whether they were satisfied with snow removal from streets in Fitchburg. One hundred and eight people said they were satisfied. Twenty-five people said they were not, and 18 refused to answer my question.

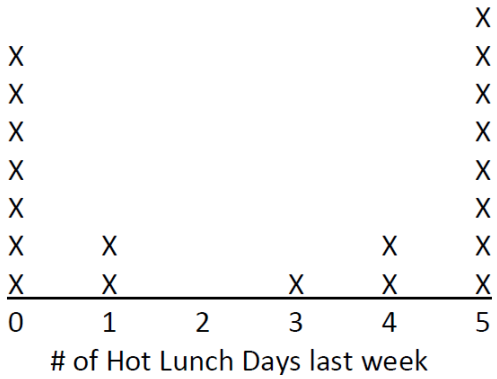
a) What is the *sampling rate?*

b) What is the *response rate?*

d) What type of bias’ may be present in this study. You may list more than one. Explain each choice.

Example 6.1.5

Suppose your third graders have collected data from their class on the number of times last week that they ate hot lunch at school and displayed it in the following so called *line plot* (each x represents 1 child in the class):



There are many questions you might ask your students about these data – but there are some questions particularly related to *sampling* that should be asked and discussed. Read each question and respond to the prompts that follow.

Question 1: Was there anything special about *last* week that might have made these data different than if we'd asked the same question next week or do you think last week was a typical week?

> *Prompt 1: In what ways might your students respond to this question? What points about sampling could you make in each case?*

Question 2: How do you think the data might have been different if we'd asked the *fifth* graders the same question?

> *Prompt 2: In what ways might they respond to that question? What points about sampling could you make in each case?*

Questions 3: Would it be okay to label this graph “Number of Days Each Week that Children Eat Hot Lunch at Our School?” Why or why not?

> *Prompt 3: In what ways might they respond to that question? What points about sampling could you make in each case?*

Notice that these questions help children think about whether the data is representative of a population larger than the sample of 20 third-graders shown on the graph, and it helps them to begin to think critically about what the data says and what it does not say.

Example 6.1.6

Consider the following claim from a company that produces a certain weight loss supplement:

“Four out of five TikTok influencers recommended this weight loss supplement for people who are overweight”

Think of some questions you could ask about the corresponding data behind this claim. Discuss.

Surveys and data can be misleading. It is important for you and your students to think critically about information based on survey data. In particular, you should consider the following:

1) Sampling method: Surveys are meant to provide information about a population, but a survey is only as accurate as its sample in representing that population. Any method other than random sampling is likely to produce a biased sample. This includes approaches that systematically remove parts of the population or those that force diversity or intentionally change the representation of socially identified groups. Such approaches create samples that may misrepresent the population, affecting the usefulness & generalizability of any data collection and analysis that follows from it.

   A random sample will tend to be representative of the population. If a population is diverse, a random sample will tend to pick up that diversity in its appropriate amounts, provided that it is not too strongly affected by a low response rate or self-selection bias.

2) Wording of Questions: Questions used by questionnaires or interviews should use neutral language and be carefully worded to avoid being misinterpreted, misleading, or otherwise influencing the responses. Also, depending on the question being asked, self-reported data might have inaccuracies that could be avoided.

3) Analysis: The information we get from reliable data from a representative sample depends upon the values we choose to calculate. In the next section, we discuss measures of central tendency (for example: average) and what they tell us (or do not tell us) about our data.

4) Presentation: Results from studies are often presented with explanations and graphs. There are many ways to present data and some of these can be misleading or inaccurate. Later in this chapter we will look at different types of graphs and learn how to read and critique them.

Done carefully, statistical studies can give us an accurate picture of a population and its characteristics. Statistics can be very helpful in making important predictions and in advising change. For example, statistics are used to make weather forecasts, to advise health and security measures. Statistics may also be used to help second grade teachers understand their students' issues and interests.

As teachers you may gather data for your students to analyze from the U.S. Census Bureau website and other sources. For example check out the the following data on bullying from the National Center for Education Statistics
https://nces.ed.gov/fastfacts/display.asp?id=719

6.1 Exercises

1) Explain the difference between a parameter and a statistic.

2) Suppose you are trying to gather information about the ethnic background of students at an elementary school. Assume the school has 500 students and you are trying to minimize sampling error. Give examples of each of the following. In doing so, state the size of your sample and be clear about exactly how your sample is selected

    a) A simple random sample

    b) A cluster sample

    c) A systematic sample

3) The Leominster City Council wants to know whether the 43,613 citizens of the city approved of a proposal to spend $1,000,000 to revitalize the downtown. The following study was conducted for this purpose. A survey was mailed to every 10th person on the list of registered voters in the city (it was mailed to 3000 people) asking about the proposal. Citizens were asked to complete the survey and return it to the Council by mail in a self-addressed, stamped envelope. 2700 people responded; of these, 75% supported the expenditure and 25% did not.

a) The population for this study is

A) all registered voters in the city
B) all citizens of the city of Leominster
C) the 3500 people who were sent a survey
D) the 3100 people who returned a survey
E) none of the above

b) The "75%" reported above is a

A) parameter
B) population
C) sample
D) statistic
E) none of the above

c) What is the sampling rate for this survey?

d) What is the response rate for this survey?

e) True or False? The results of this survey may be unreliable because registered voters may not be representative of all the citizens. Explain your answer.

f) True or False? The method of sampling in this survey could best be described as simple random sampling. Explain your answer.

g) What type of bias(es) may be present in this study. You may list more than one. Explain each choice.

4) A researcher wants to predict the outcome of the Frostbite Falls mayoral election. In order to do this, she mailed a survey to every tenth person on a list of all 40,980 registered voters in the city. One thousand fifty-six people returned the survey. 34% of respondents said they supported Snidely Whiplash; 48% said they supported Dudley DoRight; the remaining 18% were undecided.

   a) The population for this study is
      i) all mayoral-race voters in the city of Frostbite Falls
      ii) all citizens of the city of Frostbite Falls
      iii) the people who were sent a survey
      iv) the people who returned a survey
      v) none of the above

   b) What is the response rate?

   c) What is the sampling rate?

   d) The "48%" reported above is a
      i) parameter
      ii) population
      iii) sample
      iv) statistic
      v) none of the above

   e) This survey suffers primarily from
      i) sampling bias
      ii) non-response bias
      iii) sampling error
      iv) having a sample size that is too small

5) Sometimes a website or publication will advertise a poll asking its viewers to 'call in' or 'click to respond' to questions on a topic. In this case, we say that the sample is **self-selected**. What types of sample biases are likely in this type of survey? Explain.

6) A middle school is planning a long weekend class trip for their 8th graders and asked a few trusted students to help decide what location they should go to out of the following: New York City, Quebec or Washington DC. Each of the three students (Neave, Mario and Anya) decided to collect corresponding data from the 350 members of their 8th grade class and their responses were follows:

   Neave: *We should go to New York City. A couple of days last week, I checked the cafeteria during lunch and walked the halls during passing time, looking for 8th graders. To be fair, I found 30 boys and 30 girls, each from a variety of races/ethnicities. Of them, 25 preferred New York City, 15 preferred Quebec, 19 preferred Washington DC, and 1 didn't care either way.*

   Mario: *No, we should go to Quebec. I figured this out by taking a list of all 350 8th graders and rolled a standard die for each one; I decided to ask all those for whom I rolled a 6. In the end, 20 preferred Quebec, 10 preferred New York City, 7 preferred Washington DC, 1 didn't care either way, and 2 people couldn't be reached. So most want the Quebec trip.*

   Anya: *Actually, most people want to go to Washington DC. I emailed everyone in the 8th grade class to find out what they wanted. Of those that responded, 25 wanted to go to Washington DC, 11 wanted to go to New York City, 11 wanted to go to Quebec, and 3 said they had no preference.*

Based only on their arguments, where would you (as their advisor) recommend that the 8th grade trip take place, New York City, Quebec or Washington DC? Give your best argument to justify your decision.

6.2 Graphing Data

The following quote from the National Council of Teachers of Mathematics Principles and Standards for School Mathematics, p.176 provides a good motive for this section.

*Instructional programs from **prekindergarten through grade 12** should enable students to formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them.*

Graphing data is not only useful in mathematics, but may also be used in other subjects, for example science, to investigate a hypothesis by plotting the findings in a graph.

Pre-Class Work

1) Watch the video and complete the 'test yourself' exercises at the following link https://studyjams.scholastic.com/studyjams/jams/math/data-analysis/line-plots.htm
2) Watch the video and complete the 'test yourself' exercises at the following link https://studyjams.scholastic.com/studyjams/jams/math/data-analysis/bar-graphs.htm

It is important that children not only learn to display data but also learn to make sense of charts and graphs they find in books and other media and to be critical of misleading displays. Children in elementary school learn to make and interpret pictograms, frequency plots (sometimes called line or dot plots), bar graphs, and pie charts. Each type of data display has its own benefit. No matter what type of display we use, titles and labels will be very important. Let us distinguish among these types of displays through an example.

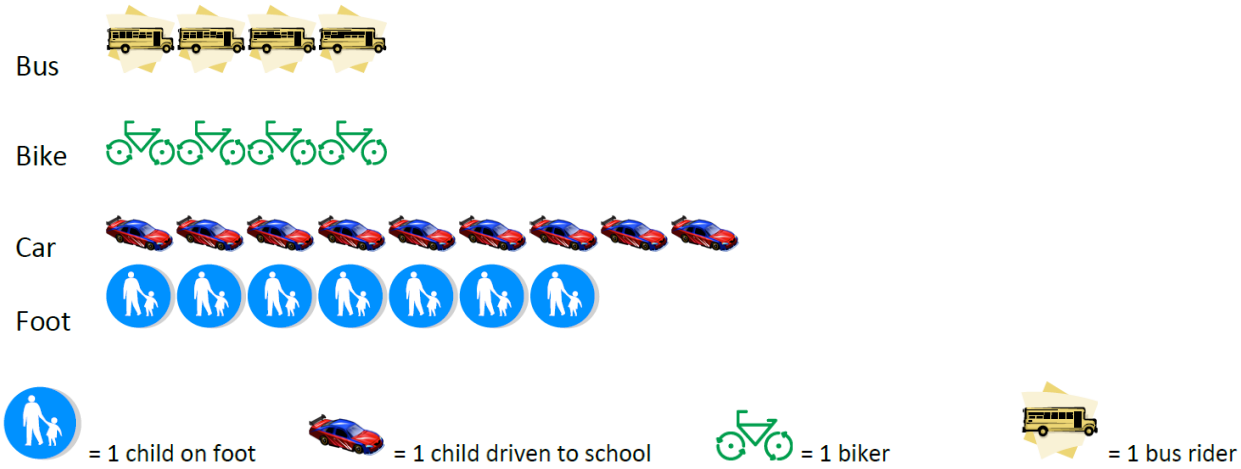
The data below show how children in Mr. Jensen's class travel to school.

Student	Transportation	Student	Transportation	Student	Transportation
Abby	Bus	Sasha	Foot	Xia	Foot
Winton	Bicycle	Lucy	Bus	Kim	Foot
Dexter	Car	Joe	Car	Ben	Car
Trevor	Car	Aiden	Bus	Leah	Car
Audrey	Car	Justin	Foot	Jeffery	Foot
Mavis	Foot	Campbell	Car	Queztal	Car
Kate	Foot	Sonja	Car	Cheyenne	Bicycle
Lim	Bicycle	Jen	Bus	Steve	Bicycle

As was illustrated in the pre-class work, representing data with pictures is a good starting point for elementary school students to organize data.

A **pictogram** or **pictograph** or **picture graph** represents each value as a meaningful icon. The display can be oriented either vertically or horizontally. We illustrate this using the transportation data as follows:

Transportation Pictogram for Mr. Jensen's Class



Notice how the graph is carefully labeled and that a key is provided for the icons. Make sure to help children get in the habit of labeling and defining their symbols. Also, stress that it is important that all the icons be about the same size or the graph could be misleading. A pictogram can be understood by children as young as those in kindergarten.

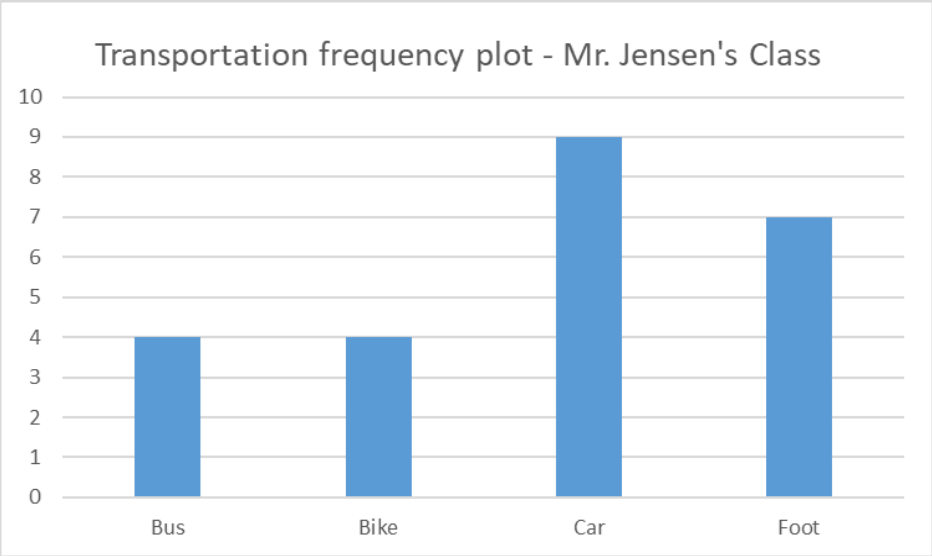
A **frequency plot** of the same data is shown below. In elementary school such graphs are often called **line plots** or **dot plots** (using dots instead of x's). Note that this plot displays the frequency of responses in a more abstract manner than does the pictogram but is easier to create.

Transportation Frequency Plot for Mr. Jensen's Class

Bus	Bike	Car	Foot
X	X	X	X
X	X	X	X
X	X	X	X
X	X	X	X
		X	X
		X	X
		X	

In a frequency plot the number of x's above each category (or value in a numerical data set) equals the number of times that category (or value) appears in the data set. A frequency plot provides a stepping stone to the next level of abstraction: a bar graph.

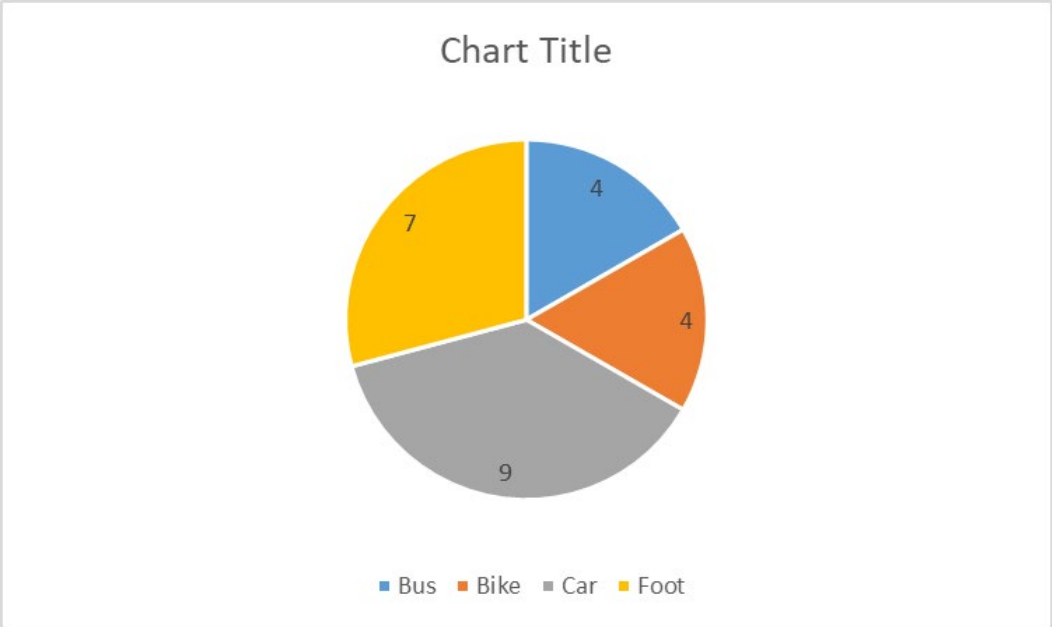
A **bar graph** uses bars (as illustrated below) to show the frequency of data in each category.



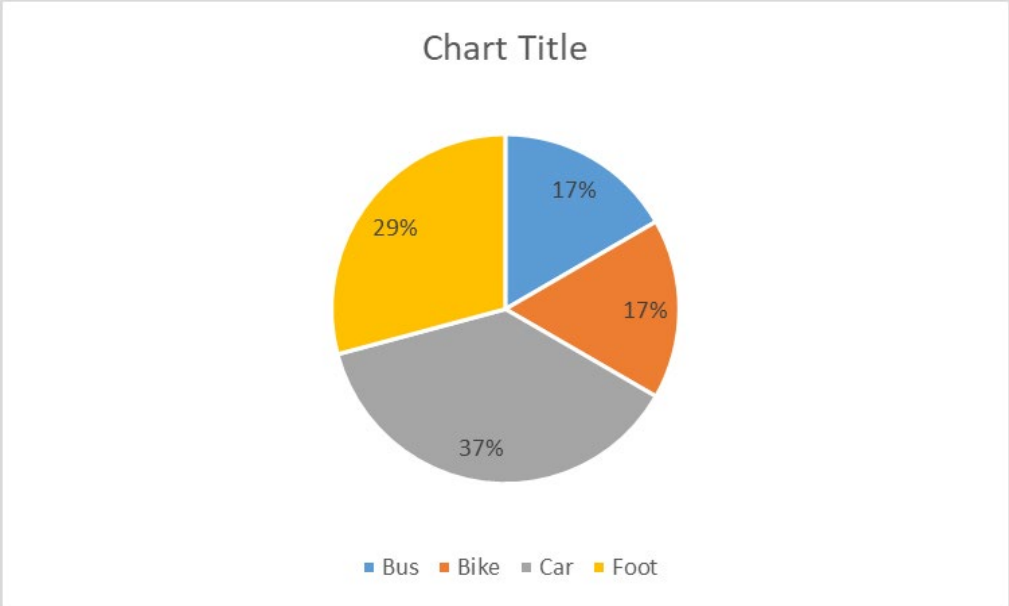
In a bar graph the height of the bar equals the frequency of the corresponding category (or value) within a set of data.

Question: Why might one want to use a bar graph instead of a frequency graph? Discuss and write your thoughts below.

Last but not least, we show a **pie chart** for the same data below.



The size of each slice in a pie chart is determined by viewing each (frequency) number as a percentage of the total and pie charts are often displayed with these percentages as shown in the following chart.



Discuss how each of the percentages in the pie chart above was calculated. Record your calculations/explanations below.

The following interactive video illustrates an example of a situation in which a pie chart is the best choice to represent the data:
https://studyjams.scholastic.com/studyjams/jams/math/data-analysis/correct-graph.htm

Example 6.2.1

Below is a table of data showing the pets owned by children in Mr. Sanchez's third grade class.

Student	Pets in the House
Anna	None
James	1 cat and 1 dog
Cleo	3 dogs
Violet	1 cat
Andre	6 fish
Jose	7 fish, 1 dog, 2 cats
Aidos	None
Yvonne	None
Tyrise	1 hamster
Ellen	2 dogs
Tomas	1 dog
Yani	3 fish and 2 cats
Owen	None
Audrey	1 cat

a) Make a frequency plot (dot/line plot) with 'number of pets' on the horizontal axis.

b) Make a bar graph of these data with 'number of pets' on the horizontal axis.

c) Make a pie chart of these data.

d) What problems might children encounter in trying to represent these data in the above ways?

Example 6.2.2

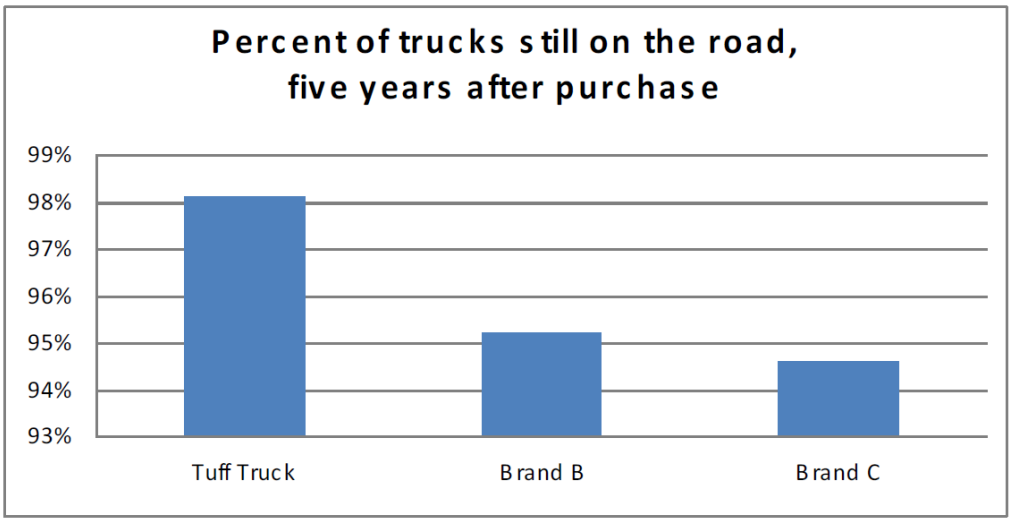
Kenya claims that the children in her kindergarten class like cats more than dogs. She has made the following graph of her classmates' pet preferences to support her claim. *Why is this pictogram misleading? Explain. What would you say to this child?*

Children who like cats 

Children who like dogs 

Example 6.2.3

Tuff Truck Company claims that their trucks are the most reliable and shows the below graph to make their case. Why is this bar graph misleading? Explain.



Stem and leaf plots

Pre-Class Work

1) Watch the following video on **stem and leaf** plots
   https://www.youtube.com/watch?v=dXwmd-98uFM
2) Consider the following stem and leaf plot representing geometry exam scores.

Exams Scores in Geometry

2 | 0
3 | 2, 7
4 |
5 | 6, 8, 9, 9
6 | 2, 4, 5, 5, 7, 8
7 | 0, 0, 2, 4, 5, 6, 7
8 | 0, 1, 1, 2, 4, 4
9 | 2, 4, 5
10| 0

a) Make a “key” for this stem and leaf plot.

b) List all the data values corresponding to this stem and leaf plot.

c) Create three questions that you could ask elementary students about this stem and leaf plot.

3) Think about how stem and leaf plots may be created for 3 or more digit numbers, for decimals, for fractions. Be prepared to discuss this in class.

A **stem and leaf plot** (or simply **stemplot**) is a representation of the data (typically) arranged by place value where each stem value is paired with corresponding leaf values. The goal of stemplots is to display data (with common place values) in a more concise way, making it easier to answer questions about the data.

Question: For what types of data sets would stem and leaf plots be impractical?

Example 6.2.4

Make a stem and leaf plot for the following. Include a key.

a) Scores from grade 4 English tests:

12, 31, 29, 52, 47, 20, 58, 43, 46, 23, 51, 14, 55, 34, 23, 35

b) A bowler's score for her 10 games in a competition:

209, 191, 218, 228, 106, 218, 213, 227, 218, 207

Line graphs

Pre-Class Work

Watch the following videos and respond to the prompts:
Line graphs (early elementary)
Line graphs (late elementary)

Prompts for pre-class work

Prompt 1: For what type of data are line graphs useful?

Prompt 2: Think of some points that are stressed in the late elementary video but not in the early elementary video. Be ready to discuss in class.

Children begin interpreting simple picture graphs from kindergarten onwards, and they are likely to study line graphs in third grade. **Line graphs** help with representing a value over time, **not to be confused with line plots which we already covered.**

Components of a line graph are as follows:

Title: This appears above the graph and should be a very concise description of the data represented.

Horizontal axis: This axis typically represents the span of dates, days, or times for which the data are collected.

Vertical axis: This axis shows the span of values collected over time.

Important notes regarding axes:

- All measurements on axes should be split equally (same scale) as stressed in the 2nd video in the pre class work.
- Each mark on the axes should be labeled with the corresponding time (horizontal axis) or value (vertical axis).
- Each axis should have a label briefly describing what the axis represents.

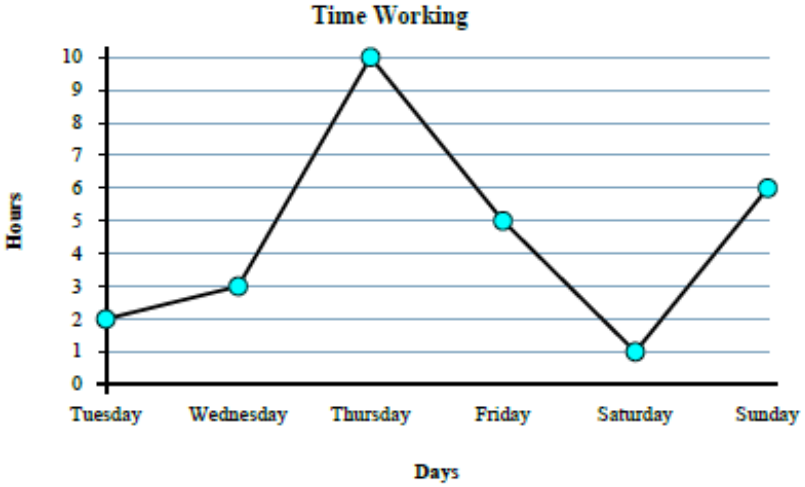
Points (ordered pairs): First number represents time and second number represents the value at that time. These are plotted on the axes as shown in the pre class work videos.

Line: Each pair of consecutive points is connected by a straight line segment.
Elementary students first learn to read line graphs before creating them.

The following example is a good type of problem to start with.

Example 6.2.5

The graph below shows the number of hours Shiloh worked each day of the week. Use the graph to answer the questions that follow.



a) How many hours did Shiloh work on Sunday?

b) Which day did they work the least?

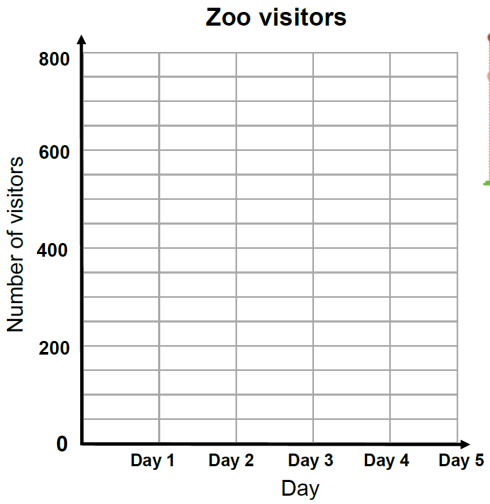
c) What is the total number of hours Shiloh worked for the week?

d) What is the difference in the number of hours they worked on Tuesday and the number of hours they worked on Friday?

When showing elementary students how to plot line graphs, we may start with examples that provide labeled axes as illustrated below.

Illustration

From K5 Learning
A zoo has a record of the number of their visitors for five days. Draw a line graph using the given data.



Day	Number of visitors
Day 1	300
Day 2	700
Day 3	800
Day 4	450
Day 5	600

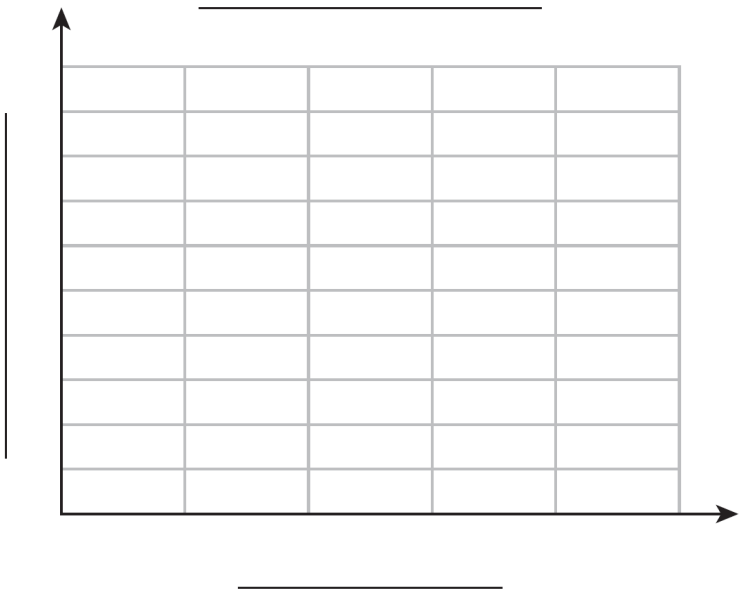
These types of problems will help prepare them to create their own line graphs in later grades.

The following example illustrates the next step toward creating line graphs. It provides the axes and places for labels but everything else is left to the student.

Example 6.2.6

The number of absences from grade 1 to 5 at a school in a month are given below. Make an appropriate scale and draw a line graph. Also label the axes and write a title for the graph.

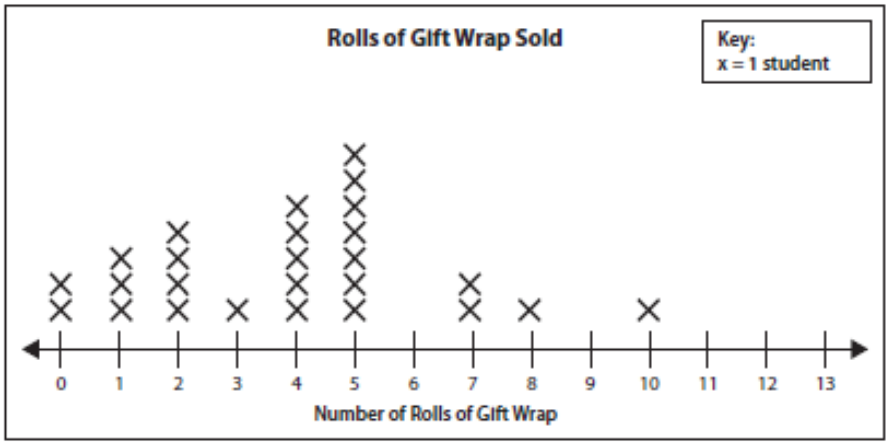
Grade	Number of Students
Grade 1	15
Grade 2	6
Grade 3	18
Grade 4	6
Grade 5	9



Extension activity: Research “double line graphs” before next class and come to class with an example to share.

6.2 Exercises

1)
Consider the following “Gift Wrap Fundraiser” problem from Bridges in Mathematics grade 3, unit 2. Answer the questions in the problem itself plus the additional “teacher questions” that follow.



a) How many students are in the class? Explain how you know.

b) How many students sold 7 rolls of gift wrap?

c) How many rolls of gift wrap did most of the students sell?

d) Sarah sold more gift wrap than anyone else. How many rolls did she sell?

e) How many rolls of gift wrap were sold in all? Show your work below.

Teacher Questions

a) Explain why a child might give the answer of “5” to question 2 about how many students sold 7 rolls of gift wrap.

b) Explain why a child might give the answer of “7” to question 4 about how many rolls of gift wrap Sarah sold.

c) Explain as if to a student, why the answer to question 5 is NOT the total number of Xs on the dot plot (line/frequency plot)

2)
Amelia recorded the amount of milk she uses in 14 different desserts (see table below). Draw a line plot (frequency plot) and answer the questions below.

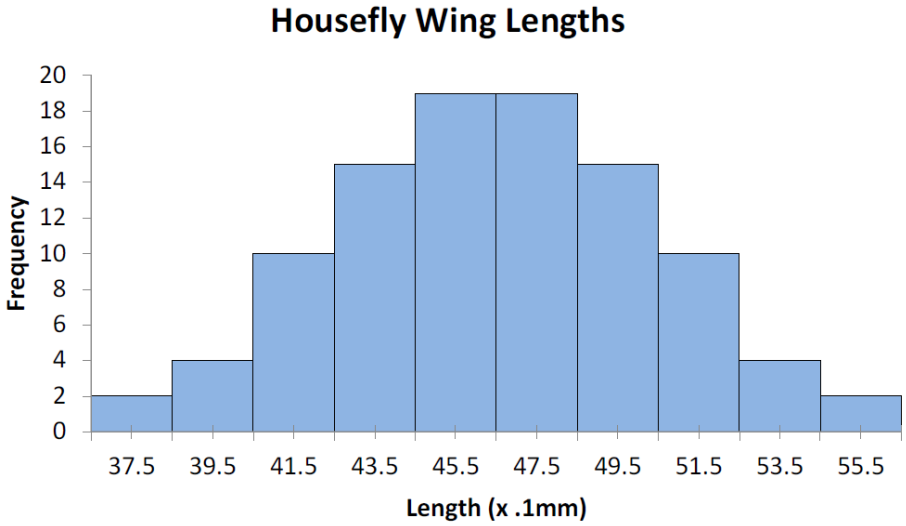
Amount of milk
1 ½ cups
2 ¼ cups
3 cups
2 ¼ cups
2 cups
2 cups
1 ½ cups
1 cups
1 ½ cups
3 cups
2 cups
2 ¼ cups
1 ½ cups
2 ¾ cups

a) How many recipes used more than 2 cups of milk?

b) How many more recipes used 2 ¼ cups than used 3 cups?

c) How many cups did she use in total? (do this without a calculator)

3)
This graph shows lengths of the wings of houseflies from the *Quantitative Environmental Learning Project.* (Original data from Sokal, R.R. and P.E. Hunter. 1955. A morphometric analysis of DDT-resistant and non-resistant housefly strains *Ann. Entomol. Soc. Amer.* **48:** 499-507)



a) Approximately how many flies were measured for this study? Show and explain your work below.

b) About what percent of the flies have wing length smaller than 5 mm. Show and explain your work below.

c) Make up a good question to ask an elementary school child about this data set.

4)

Watch the following video to learn how to use excel to create bar graphs, pie charts, and line graphs: Creating pie, bar and line charts. Then, use excel to create such graphs using data from one or more of the problems in this section. Be prepared to share your graphs in class.

5)
The ages of employees at a bank are listed below. Make a stem and leaf plot (with key) of the data.

22, 31, 43, 33, 51, 25, 37, 45, 29, 38, 41, 25, 44, 28, 50, 32, 25

6)
Dan got a job selling smoothies during his summer vacation. His earnings (in dollars) each day are recorded in the stem and leaf plot below.

Stem	Leaf
0	5 8 9
1	0 1 2 2 2 7
2	1 2 3 6
3	0 1 5

Key: 2 | 3 = $ 23

a) How much money did he earn on his best day?

b) How many days did he earn less than 20 dollars?

7)
The stem and leaf plot below shows the number of pages of a book read by each student in an hour.

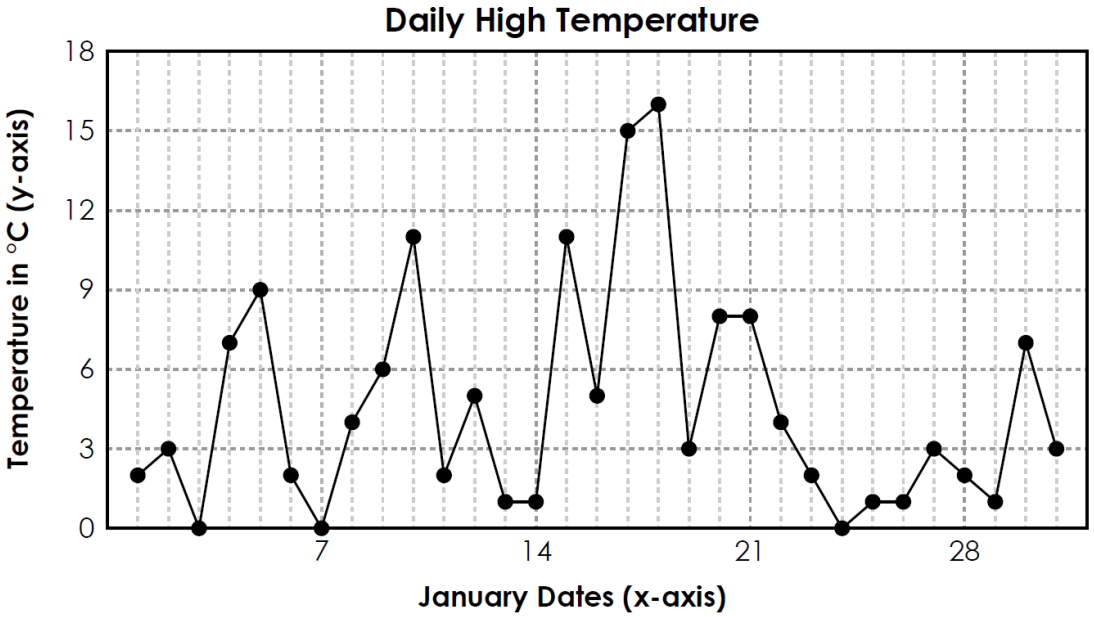
Stem	Leaf
1	3 4 6 8
2	0 1 3 3 6 8
3	1 3

Key: 2 | 6 = 26 pages

a) How many students are in the plot?

b) How many students read the same number of pages?

8)
The daily high temperatures for Gotham City in the month of January were recorded and graphed below. Use the graph to answer the questions that follow.



a) What was the approximate high temperature in Gotham City on January 10?

b) On which days was Gotham City's high temperature 7 degrees?

c) On which day did Gotham have the highest temperature?

9)
Create a line graph for the following data on Houda's sleep schedule over one week.

Day	**Hours of sleep**
Monday	7.2
Tuesday	6.5
Wednesday	6
Thursday	8.4
Friday	9.2
Saturday	9.1
Sunday	5.8

10)
Create 3 types of line graph problems for elementary students as follows.

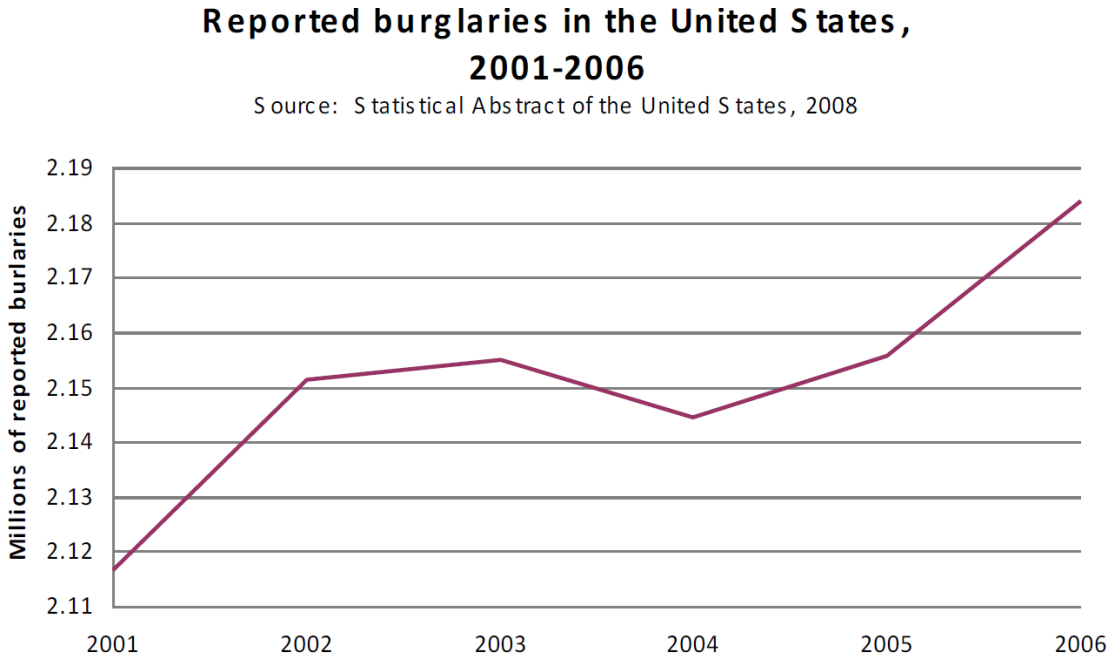
a) Given a line graph and 3 questions.

b) Given data and labeled axes for them to graph on.

c) Given data for them to make a line graph for.

11)

Consider the line graph below that was used to claim that there was a national 'crime wave' during the given time period. Does the graph provide compelling evidence of this claim? Explain why or why not.



Extra (pedagogical) exercises

1) Review the following lesson plan from Bridges in Mathematics
   Data Analysis: Bar Graphs (grade 1)
   Create a similar activity appropriate for 1st grade students. It should include materials needed and detailed instructions. Use a separate piece of paper.

1) Review the following lesson from Bridges in Mathematics
   Analysis: Graphing (grade 3)
   Complete Independent Worksheets 1 and 2

2) Review the following lesson plan from Bridges in Mathematics
   Measurement and Data: Line Plots (grade 3)
   Create a similar activity appropriate for 1st grade students. It should include materials needed and detailed instructions. Use a separate piece of paper.

3) Review the following lesson from Bridges in Mathematics
   Data Analysis: Line Plots (grade 4)
   Complete Independent Worksheets 1 and 2

6.3 Measures of Central Tendency

There are three statistics that are commonly used to describe the center of a distribution of numerical data: mean, median and mode.

The **mean** of a set of data is the average of all the values. That is, the mean is computed by summing the values and dividing by the number of values in the data set.

The **median** of a set of data is the middle value (assuming the data is in numerical order). If there are two middle values, the median is the average of the two middle values.

The **mode** is the most frequently appearing number among the data values. There may be more than one mode.

First we focus on mean, and later we will address median and mode.

Children in grades 3-5 can easily understand and compute median and mode, but typically the idea of mean is a more difficult topic. This is not because it is hard to calculate a mean but rather because it is more difficult to understand the idea of a mean.

Before showing elementary school students the standard computation of the mean, teachers should help them understand the concept which can be done by addressing the mean as a *redistribution* process or as finding the *balance point*.

The redistribution concept of the mean: The mean value of a set of numbers is what each data point would get if the data was redistributed so that each data point had the same amount.

This concept is essentially the same as the partitive concept of division: taking a total amount and sharing it equally into a number of groups, and seeing how much each group gets.

Another way to look at mean as redistribution is “evening off” the data. That is, the mean is the value each person would get if all the data was shared evenly.

For example, suppose the picture below shows the cookies that belong to each child.

Keesha's cookies: 

Andi's cookies: 

Tim's cookies: 

Myosia's cookie: 

Carlos' cookie: 

In order to find the mean number of cookies, the children could imagine sharing cookies (or even parts of cookies) so that each child has exactly the same number of cookies as follows.

Keesha's cookies: 

Andi's cookies: 

Tim's cookies: 

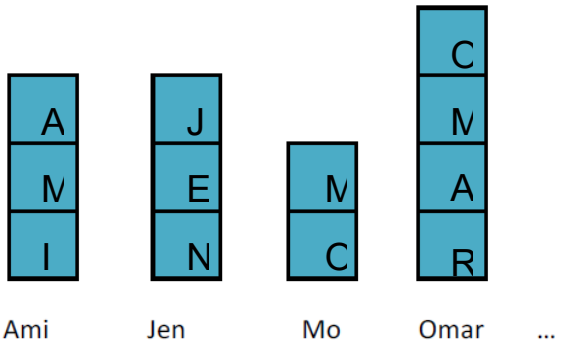
Myosia's cookies: 

Carlos' cookies: 

So, the children see that the mean number of cookies is 3. In this case we didn't need to break up cookies into pieces (fractions of a cookie) in order to share them, but you should have such examples in your future classes.

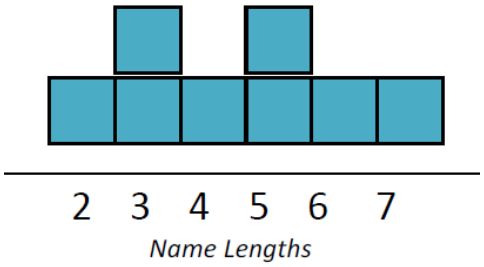
Class Activity

a) Each person in the class write his/her/their first name on sticky notes so that there is one letter on each sticky note. For example someone with the first name “JEN” would need 3 sticky notes to write the letters J, E and N on each one respectively. Line the sticky notes up on the board in vertical stacks/piles as illustrated below.



Proceed as follows to find the mean/average number of letters in first names in the class: Rearrange the sticky notes so that every stack/pile ends up with the same number of letters/sticky notes or equal parts of sticky notes. Use scissors to cut sticky notes in parts as needed to get the piles exactly equal in height.
Write the mean you found below.

b) Now organize the first name data in another way as follows: Each person gets ONE sticky note, writes their entire first name on it, and places it on a so-called “bar graph” on the board, based on the number of letters in your first name. For example, if your first name is “JEN” then you would place your sticky note above the number 3. For clarification, an illustration is shown below for names Mo (2 letters), Ami (3 letters), Jen (3 letters), Omar (4 letters), Karen (5 letters), Skyla (5 letters), Jethro (6 letters) and Juanita (7 letters).



Once all the sticky notes are on the board, use the bar graph data to help calculate the mean name length for the class. Show your work below.

Example 6.3.1

The family sizes for children in a class are shown below. Make a bar graph for these data and then find the mean number of family members.

Family	size
Jenson	8
Lewis	3
McDuff	7
Earles	4
Seaman	5
Ortiz	4
Peters	6
Todd	5
Miene	3

Explain how you could use the redistribution concept of the mean to solve this problem. Show and explain your process below.

The balance point concept of the mean: Viewing our data values as equal weights placed on a ruler or number line level in a location that corresponds to their value, the balance point would be where we have to put a pivot in order for the level to balance. The image below illustrates some trial and error in finding the mean (as a balance point) of the data set 1, 2, 6, 7, 9.

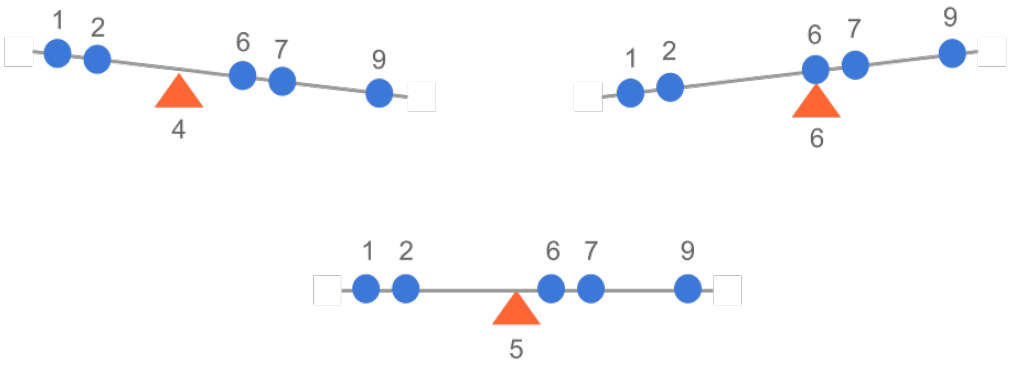


Image from https://www.gastonsanchez.com/matrix4sl/mean-as-a-balancing-point.html

Pre-Class Work

How to find mean as a balance point

1) Read and work through examples in the following lesson: mean as the balancing point

2) Watch the following videos:
   Balance point concept of the mean
   Was there anything unclear in the presenter's words? Be prepared to discuss.
   Balance point missing value
   Explain below in your own words how distances relate to the mean of a data set.

Let us revisit the cookie example. Recall the number of cookies for each student gives us the following numbers: 4, 7, 2, 1, 1. Use the method shown in the (balance point) videos to find the mean. Show your work below.

Now check if the following equality (as claimed in the second video) holds for the cookie data:

Sum of the distances from each data point to the left of the mean, to the mean value

EQUALS

Sum of the distances from each data point to the right of the mean, to the mean value

Show your work below.

Question: In what type of scenario would the line/frequency plot technique used in the pre-class work videos NOT work to find the mean? Discuss.

Example 6.3.2

Suppose there are 4 cats with a mean weight of 10 pounds. You know the weights of 3 of the cats. Those weights are: 4 pounds, 7 pounds, and 15 pounds .

a) Use the equality below to find the weight of the 4th cat.

Sum of the distances from each data point to the left of the mean, to the mean value
EQUALS
Sum of the distances from each data point to the right of the mean, to the mean value

Show your work below.

b) Check your answer for part (a) by finding the average weight of all 4 cats. That is, by adding all the weights and dividing by 4.

As illustrated by the previous examples, the *balance point concept of the mean* views the mean as the value such that the sum of the distances that each data point is above this value balances with the sum of the distances that each data point is below this value.

Now let us try to solve some slightly more advanced mean problems using arithmetic and/or algebra.

Example 6.3.3

You have taken four math quizzes. Your scores were 82, 67, 85, and 72. What do you need to score on the 5th quiz for your average quiz score to be 75? One way to solve this problem is to let a variable represent the number you are looking for, set up an equality with this variable using the definition of the mean and solve the equation. However, feel free to use another method.

Example 6.3.4

Last semester I had two sections of Math 1600 that took an exam. Section 1 had 25 students, and the mean score was 65 points. Section 2 had 15 students and the mean score was 75 points. Determine the mean score over all 40 students.

So far we have been dealing with mainly whole number data values. It is important in the elementary classroom to include activities with fractions as well to keep students comfortable and confident with all types of numbers.

Homework/Classwork Activity

Before class:

1) Watch the following video which illustrates mean in an application involving fractions https://www.youtube.com/watch?v=DtWovvMnPrk .
2) Make up your own such problem (involving mean and fractional data). Write it on a piece of paper

During class:

3) Swap problems with a classmate and solve each other's problems (show all work).
4) Swap back to correct each other's solutions and give feedback.

Now we address other measures of center (aside from mean): **median** and **mode**.

The **median** of a set of data is the middle value (assuming the data is in numerical order). **If there are two middle values** then we average these two to get the median of the data set.

Once again let us recall the cookie problem: The number of cookies held by the five people is 4, 7, 2, 1, and 1. If we put these values in numerical order, we get: 1, 1, 2, 4, 7

We see that the middle value is 2 (we'll write M = 2.).

Why should we put them in numerical order first? How would you explain this to a child?

In the case of an even number of values, for example 1, 1, 2, 3, 4, 7, we have no single middle value, and so the median is the average of the two middle values 2 and 3. That is, the median of this data is 2.5.

The **mode** is the most frequently appearing number among the data values. There may be several modes in a single data set if there are values with the same maximum frequency.

What is the mode for the cookie data? 1, 1, 2, 4, 7

Example 6.3.5

Find the mean, mode, and median of bowling scores below:

112 114 120 123 127 127 130 167 220

Mean:

Median:

Mode:

Why does it make sense that the mean is larger than the median for these data? Explain.

Example 6.3.6

A data set consists of 10 whole numbers. The median is 4 and the mean is 5. What are the largest and smallest possible values for the maximum? Hint: Think of the numbers in ascending order so the maximum number is last, then determine what the mean and median tell us about the data. Don't be afraid of some trial and error. Show your work below.

a) Largest maximum value:

b) Smallest maximum value:

Class Activity

Go to https://digitalfirst.bfwpub.com/stats_applet/stats_applet_6_meanmed.html
Read how the app works and play around with it to get a better understanding.

Then, proceed as follows: Create three data sets, all of which have 6 data points and a mean and median of approximately 50 that meet the following criteria. For each set, make a sketch of the corresponding number line with your data points below. The data point values are not labeled in the app so just do the best you can with approximating where the points should be based on the criteria

- Set A: Every data point is between 35 and 65.

- Set B: Every data point is either less than 25 or greater than 75.

- Set C: The difference between every pair of two consecutive data points is the same.

Respond to the following prompts based on your exploration with the Mean and Median app.

1) Discuss the similarities and differences between the data sets. Record notes below.

2) Are there other data sets with 6 points, a mean of 50, and a median of 50 that look different from the three you've created? If so, give an example below.

3) What did you learn about mean and median by doing the exploration provided using the applet?

4) How might these ideas help students understand how the mean and median relate to a distribution of data?

It is important to know which measure of central tendency (mean, median or mode) is “better” to use in a given situation. This depends on which type of questions you are considering. The mode, for instance, can be used for categorical data, but the mean and the median do not make sense in that context. **Why?**

If you wanted to know how much money you would make working at Microsoft Corporation, would it be more honest to look at the mean income of its employees, or the median income? **Explain.**

Now we address a concept that is closely related to the mean and important for any teacher to understand.

Weighted average

When all the grades in a class are weighted equally, in order to get the final average we simply add up all the grades and divide by the number of grades, that is we find the mean or average of all the grades. However, more often than not, some grades are weighted differently (that is worth more or less) than other grades in the class. In this case we need the concept of *weighted average* to calculate overall averages.

Weighted averages are also used in many other applications related to data analytics, finance, and more.

Pre-Class Work

Weighted Average

1) Watch the following videos to learn about *weighted average*.
   How to calculate grades using weighted average

   Using excel to calculate grades

2) Explain below in your own words how to calculate weighted averages.

In summary, the weighted average of a set of values is obtained as follows:

1) Multiply each value in the set by its corresponding weight (in decimal form)
2) Add up the products from step (1)
3) Divide the sum from step (2) by the sum of all weights (in decimal form).

Class Activity

In class in groups
Complete the following exercises related to weighted averages

1) A certain Algebra I class has weighted grades. The weights are:
   10% Homework
   30% Quizzes
   60% Tests

A grade sheet is given below. Calculate the student's current average based on these grades.

Assignment	Grade
HW 1	100
HW 2	100
HW 3	105
HW 4	96
HW 5	97
HW 6	100
Quiz 1	68
Quiz 2	84
Test 1	78

Important Note: In terms of grading, if all grades are in and a final average is being calculated then the sum of all weights (in decimal form) should be 1 (corresponding to 100% of the work being considered), as in #1 above. However, there are some situations in which the sum of all the weights is not 1. For example when you want to calculate a student's average before all the grades are in, at which time only some of the work is being considered. In #2 we see an example of this situation.

2) Lina has the grades below so far in her math class in which homework is worth 30% of the final average, quizzes are worth 30% and midterm and final exams are worth 20% each.

Assessment	**Grade**
HW 1	95%
HW 2	90%
HW 3	92%
Quiz 1	80%
Quiz 2	70%

a) Calculate Lina's current average in the class.

b) Assuming Lina's homework and quiz averages remain consistent, what will she need to average on her exam scores in order to get an A (90% or higher) in the class?

3) Describe below the difference between "average" and "weighted average" as if you were trying to explain it to an elementary school student.

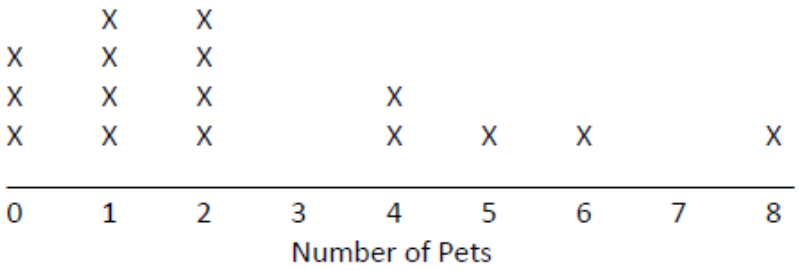
4) Tabitha ordered dinner for a party of 10 people. Three people ordered the $7.75 chicken dinner, two people ordered the $10.95 fish dinner and five people ordered the $8.75 beef dinner. What was the average cost of each dinner at Tabitha's party? Calculate this in two ways as follows. Show your work.

    a) Using weighted average

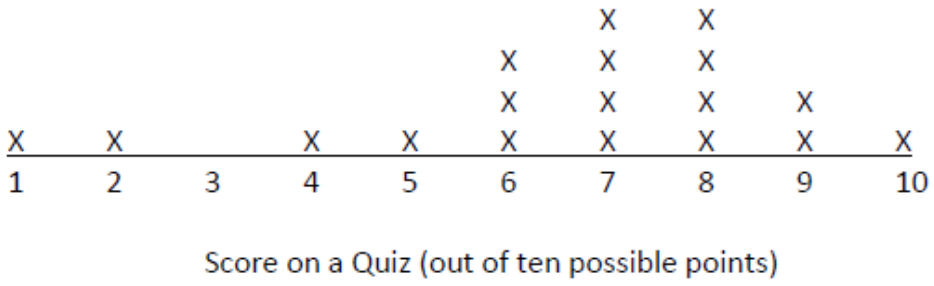
    b) Not using weighted average (just using regular average/mean)

6.3 Exercises

1) Below is a line/frequency plot showing the number of pets in 16 households. Find the mean number of pets per household. Show your work.



2) Consider the frequency plot below of 18 scores on a quiz.



a) Find the mean score.

b) Find the median score and explain your process below, as you would to a student.

c) Why does it make sense that the mean is smaller than the median for these data? Explain below.

3) Consider the following graph of the number of people in 6 families.

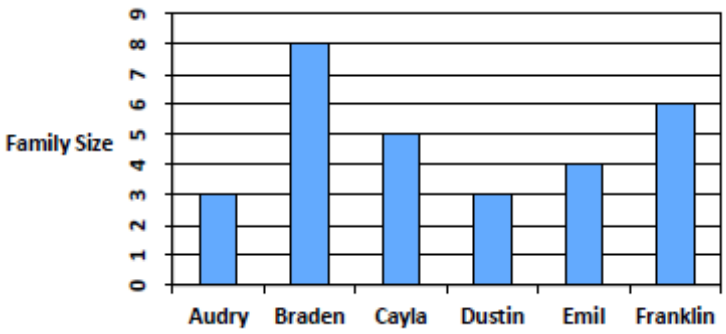


Figure out how you can use the graph to find the mean family size, without doing any calculations (hint: use redistribution). Explain how you did it.

4) Ten kids were served cookies and they were eaten as follows: Three kids ate 1 cookie each, two kids ate 2 cookies each, two kids ate 3 cookies each, two kids ate 5 cookies each and one kid ate 7 cookies.

   a) Sketch the data on a line/frequency plot

   b) Find the mean number of cookies eaten (per kid).

5) Here is a data set of the number of years members of the math department have been employed at a school: 7, 25, 7, 15, 8, 23, 27, 6, 1, 3, 28, 22, 24, 9, 15, 14, 18, 8. Compute the mean, median and mode for these data. Show your work.

   Mean:

   Median:

   Mode:

6) Decide whether each of the following is True or False. Explain your answers.
    a) If ten people took an exam, then it is possible that all but one of them scored less than the mean.

    b) If ten people took an exam, then it is possible that all but one of them scored less than the median.

    c) If ten people took a 100-point exam, it is possible that the mean and median are 90 points apart.

7) At a job fair, representatives from three different departments approached you, trying to recruit you to their program. When asked how much their graduates make, here were their responses:
    a) Representative 1: Ten people graduated recently with our degree and their mean annual income is $75,000.
    b) Representative 2: Ten people graduated recently with our degree and their median annual income is $75,000.
    c) Representative 3: We have recently graduated one hundred people and their mean annual income is $75,000.

    Which program would you be most inclined to choose, if money were your only concern? Which would least interest you? Explain.

8) Go to the following link: https://www.statcrunch.com/applets/type1&meanmedian . Read the instructions. Do the “Try” activities (copied below) and put your answers to the corresponding prompts below.

   a) Create a set of values very close together. Then add a value far away from this set. What’s the impact of the far away value on the mean and the median?

   b) Create a data set with a median below 2 and a mean between 2 and 4. Copy our data set below.

   c) Specify a sample size and then select Simulate to generate a variety of data sets. Make observations about the mean and median in relation to your simulations and record your observations below.

9) Ann needs a mean of 80 on her five exams in order to earn a B in her class. Her exam scores so far are 78, 90, 64 and 83. What does she need to get on her fifth exam to earn the B?

10) Results for a 20-point multiple choice math quiz are as follows: 9 students scored a 20, 4 students scored 19, 3 students scored 18, 5 students scored 17, and 1 student scored 12. Figure out the class average in two ways as follows:

   a) Using weighted average

   b) Not using weighted average (just using regular average/mean)

11) Watch the video on how to calculate college GPA (grade point average) for a particular semester. Pause the video as prompted (at approximately 3:45 time) to try the example before viewing the solution. Then check your GPA for last semester and show your work below.

12) Barack is in a math class where quizzes are worth 15% of the grade, exam 1 and exam 2 are worth 20% each, and the final exam is worth 25% of the final grade. Given his grades below, compute the following.

	Q1	Q2	Q3	Q4	Q5	hw1	hw2	hw3	hw4	exam1	exam2	Final exam
score	5	6	8	8	9	10	10	0	8	41	48	89
out of	9	9	9	9	9	10	10	10	10	53	56	92

a) Barack's final average as a percent.

b) Barack's final average as a percent if the zero on HW 3 is dropped.

13) Assume you are in a class where quizzes are 20% of your grade, homework is 20%, exam 1 and exam 2 are 15% each, and the final exam is 30%. Suppose it is week 5 and you want to calculate your current average. The grades you have so far are as follows:

First three quizzes: 6/9, 8/10, 9/9
First three homework assignments: 9/11, 10/10, 4.5/7
Exam 1: 58/63

What is your current average in the course?

Extra Practice with Mean and Weighted Average (worksheet)

Chapter 6 Wrap Up

This chapter introduced statistics and how it is taught in the elementary classroom. It also addressed the importance of being able to read and analyze data, including asking critical questions as to not be misled by data that are presented with bias. Different ways to represent data were explored as well as measures of central tendency and implications thereof. This is just the beginning of the very complex field of statistics but provides a basis of understanding data and how they are presented. In middle school, students expand on these topics as well as study the spread of data and corresponding graphical representations (for example so called “box and whisker plots”). This all provides a foundation to understanding data and how they can provide valuable information (or potentially harmful misinformation). Statistics are used absolutely everywhere in today’s society and are more important than ever for students to understand.

For material on **Measuring the Spread of Data** see p. 181 - 182, 187 - 196 in Big Ideas in Geometry and Data Analysis (see link in Back Matter - OER Resources)