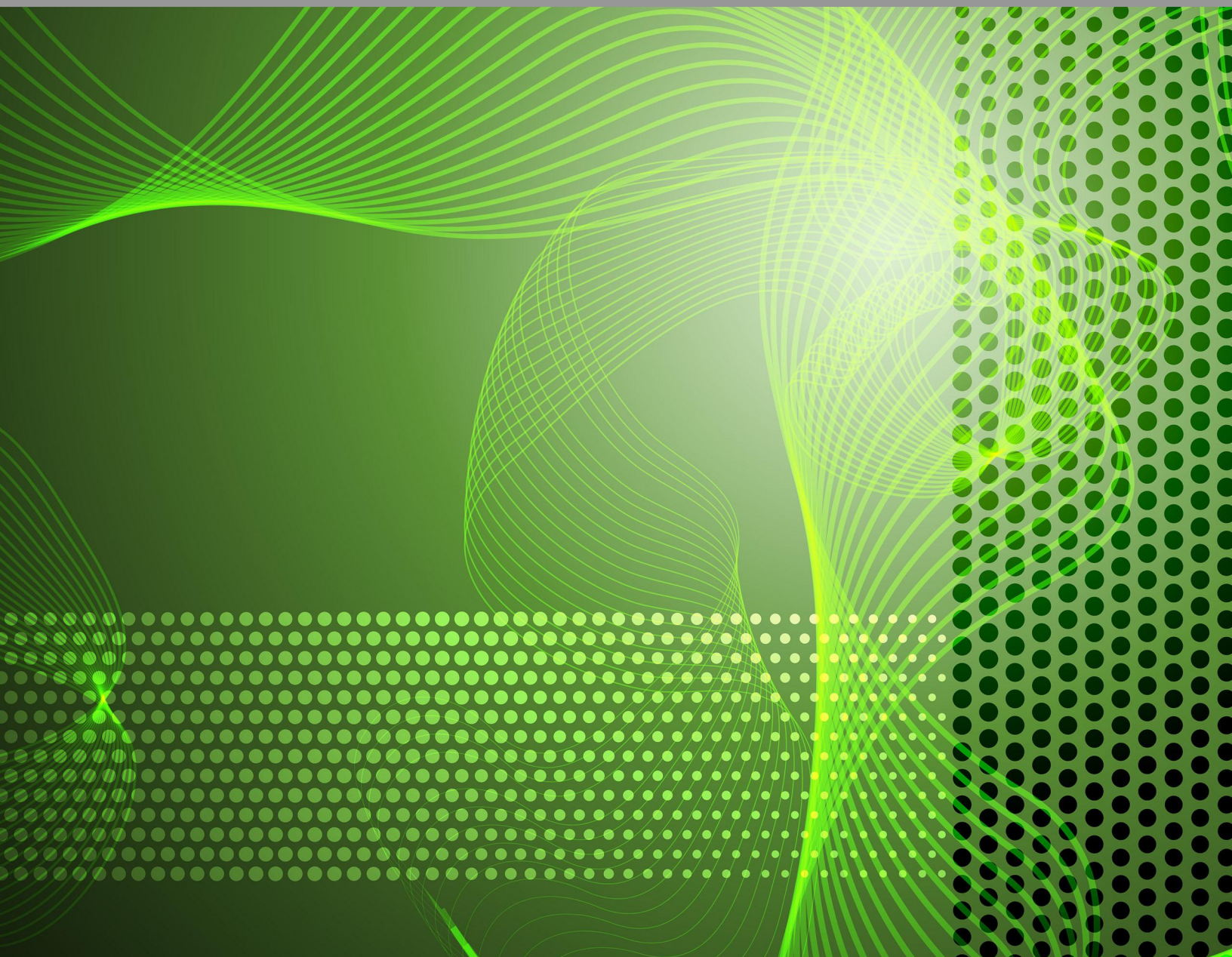


ck-12

flexbook  
next generation textbooks

# CK-12 Probability and Statistics

*Concept Collection*



# CK-12 Probability and Statistics Concepts

---

Larame Spence

**Say Thanks to the Authors**

Click <http://www.ck12.org/saythanks>

*(No sign in required)*



## CHAPTER

**4****Evaluating and Displaying  
Data****Chapter Outline**

---

- 4.1 GROUPING DATA**
  - 4.2 ANALYZING DATA**
  - 4.3 RELATIVE FREQUENCIES**
  - 4.4 CUMULATIVE FREQUENCIES**
  - 4.5 CREATING HISTOGRAMS**
  - 4.6 INTERPRETING HISTOGRAMS**
  - 4.7 FREQUENCY POLYGONS - PROBABILITY AND STATISTICS**
  - 4.8 CREATING BOX-AND-WHISKER PLOTS**
  - 4.9 INTERPRETING BOX-AND-WHISKER PLOTS**
  - 4.10 CREATING STEM-AND-LEAF DIAGRAMS**
  - 4.11 INTERPRETING STEM-AND-LEAF PLOTS**
  - 4.12 CREATING SCATTER PLOTS AND LINE GRAPHS**
  - 4.13 INTERPRETING SCATTER PLOTS AND LINE GRAPHS**
  - 4.14 CREATING PIE CHARTS**
  - 4.15 INTERPRETING PIE CHARTS**
  - 4.16 REFERENCES**
- 

Properly collecting and organizing data is important, but without a way to show the results to others in a meaningful format, data becomes much less useful. Visualizing data in the form of an image allows you to identify trends and comparisons that may be difficult to see in a list of values. Graphs, plots, and charts are the primary methods of data imaging.

Graphs and charts come in a bewildering array of shapes, sizes, and types, but there are a number of them that are used quite regularly and that form the basis for many others.

In this chapter, you will learn how to construct and interpret many common data visualizations.

## 4.1 Grouping Data

### Objective

Here you will learn about different methods of organizing data for use in statistical research.

### Concept

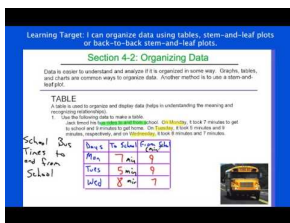
Suppose you were given the following data:

87, 72, 91, 91, 73, 83, 79, 81, 87, 72, 81, 91, 73, 73, 73

If you were told you were going to evaluate this data using common methods of central tendency and dispersion, how might you start by organizing the data in order to make the study as straightforward as possible?

### Watch This

This video presents a couple of methods of organizing data:



### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63578>

<http://youtu.be/NASBOFGMy2c?t=40s> FortLucasMath - Organizing Data

### Guidance

Data in its original form, just a list of numbers, names, letters, colors, etc., is known as **raw data**, and is often not particularly useful without some kind of organization. The series of numbers in the concept question above, for instance, doesn't really mean anything at the moment. Without some sort of context and some level of organization, this is just a bunch of meaningless values.

Data can be classified into two general types, **quantitative** and **qualitative**. There are a number of ways to group or organize each type of data to make it more useful.





- **Quantitative** Data (data that may be conveniently described numerically):
  - Dates or times are commonly organized chronologically
  - Data with recurring values is often organized in a **frequency distribution**
  - **Univariate** data likely to be used for evaluating the mean or the range of a population is generally organized in increasing or decreasing magnitude or alphabetically.
  - **Bivariate** data is usually organized in a table showing how the two variables change in relation to each other.
  - Compare and contrast tables are excellent for evaluating two or more variables
- **Qualitative** Data (data that may be difficult to describe with numerical values)
  - Commonly grouped by category
  - Categories are often evaluated using a frequency distribution
  - Data may be organized in order of importance
  - **Inductive** organization orders information by increasing complexity, listing facts prior to conclusions and advancing from specific examples to general conclusions.
  - **Deductive** organization is the inverse of inductive, listing recommendations/conclusions followed by supporting facts/data.

### Example A

Elaina is preparing to create a **histogram** to illustrate the data that she collected on average time spent taking a particular test in her Statistics class.

16 mins, 18.5 mins, 14.5 mins, 16 mins, 19 mins, 18 mins, 16.5 mins, 15 mins, 15 mins, 14.5 mins, 14 mins, 16 mins, 12.5 mins, 19.5 mins, 14 mins, 15 mins, 16.5 mins, 14 mins, 18 mins, 16 mins

- a. How should she organize the data to make the construction of the histogram as straightforward as possible?
- b. What will the data look like after it is organized?

**Solution:** A histogram is a graph that illustrates the **relative frequency** or **probability density** of a single variable.

- a. Since Elaina will need to identify the number of values in each category of the data, it would be ideal to organize the data in groups called classes or intervals. With the given data, intervals of 1 minute each would seem appropriate.
- b. Minutes Required to Complete the Test:

12.5 | 14, 14, 14, 14.5, 14.5 | 15, 15, 15 | 16, 16, 16, 16, 16.5, 16.5 | 18, 18, 18.5 | 19, 19.5

**Example B**

Orlando is planning to create a box-and-whisker plot to illustrate how much more popular dogs and cats are as pets than fish-tanks, reptiles, and birds. He has collected the data below from a randomized sample of homes in his town, using a survey questioning the number of pets each family has in each category

House 1: 2 dogs, 2 cats, 0 birds, 0 reptiles, 1 fish tank

House 2: 3 dogs, 2 cats, 1 birds, 0 reptiles, 0 fish tank

House 3: 0 dogs, 3 cats, 0 birds, 1 reptiles, 1 fish tank

House 4: 2 dogs, 1 cats, 2 birds, 0 reptiles, 1 fish tank

House 5: 2 dogs, 1 cats, 0 birds, 0 reptiles, 0 fish tank

House 6: 2 dogs, 2 cats, 0 birds, 0 reptiles, 0 fish tank

House 7: 3 dogs, 1 cats, 1 birds, 0 reptiles, 2 fish tank

House 8: 3 dogs, 2 cats, 0 birds, 0 reptiles, 1 fish tank

House 9: 2 dogs, 3 cats, 0 birds, 0 reptiles, 0 fish tank

House 10: 1 dogs, 3 cats, 0 birds, 0 reptiles, 0 fish tanks

How should Orlando organize the raw data to facilitate the creation of his box-and-whisker plot? What will the organized data look like?

Solution:

- a. Since Orlando's box-and-whisker plot is specifically meant to highlight the number of dogs and cats, it would be a good idea to organize the data in groups by importance, with dogs and cats first. Since he will need to identify the mean, range, and quartiles of the data, it would also be good to organize each group by increasing values.
- b. Dogs: 0, 1, 2, 2, 2, 2, 2, 3, 3, 3
  - a. Cats: 1, 1, 1, 2, 2, 2, 2, 3, 3, 3
  - b. Birds: 0, 0, 0, 0, 0, 0, 0, 1, 1, 2
  - c. Reptiles: 0, 0, 0, 0, 0, 0, 0, 0, 0, 1
  - d. Fish Tanks: 0, 0, 0, 0, 0, 1, 1, 1, 1, 2

**Example C**

Cheng is interested in the phenomenon of the changes in how fast time seems to pass to people as they age. He has collected data from 300 people between the ages of 10 and 70. Each person reported the time it *seemed* to take to complete three neutral (neither particularly liked nor disliked) activities, one 5mins, one 15mins, and one 60mins long. Now Cheng has a massive and somewhat intimidating list of numbers, and he needs to decide how to organize what he has into something useful.

- a. Identify at least 2 different ways that Cheng might organize the raw data that would illustrate changes in time perception as people age.
- b. How might Cheng consolidate the data so he doesn't end up needing to plot nearly 1000 values on a chart or graph?

**Solution:** With such a huge amount of raw data, Cheng's greatest challenge will be consolidating it into a useful and informative format.

- a. Cheng might choose to organize the data by increasing time in several age groups, sorting the values first by age, and then by perceived time for each activity. He might also wish to sort first by actual activity length, then by age or perceived time passage.

- b. Finding the mean perceived length for each of several age ranges would be a great way for Cheng to maintain the general integrity of his data while reducing the sheer volume.

### Concept Problem Revisited

87, 72, 91, 91, 73, 83, 79, 81, 87, 72, 81, 91, 73, 73, 73

*If you were told you were going to evaluate this data using common methods of central tendency and dispersion, what sort of preparation could you do in order to make the study as straightforward as possible?*

**Central tendency** measurements are generally facilitated by organizing data in increasing value from left to right. Ideally, it would be convenient to also note the total number of values, along with their sum, as you are ordering them.

### Vocabulary

**Central tendency** is the behavior of the 'main portion' of a set of data; the most common measure of central tendency is the mean or average.

**Probability density** is similar to **relative frequency** in that both are ways to evaluate how often a particular value or range appears in a set of data.

A **frequency distribution** is a depiction of the number of occurrences of each data point in the set.

A **histogram** is a visual way to show data, primarily used for continuous variables, that is much like a bar chart.

**Deductive** means that a conclusion was stated first, and then supported by statistics. **Inductive** means that statistical information was evaluated, and then used to postulate a generalization.

**Bivariate** data has two variables, whereas **univariate** data only has one.

**Quantitative** data is composed of numerical values, and **qualitative** data is generally not. Qualitative data is generally applied to "which of these..." questions, whereas quantitative data is commonly applied to "how many of these..." questions.

### Guided Practice

A class of 40 students took a science exam. They earned the following percentages on their tests:

73, 45, 62, 34, 59, 20, 48, 50, 78, 38, 52, 91, 57, 82, 46, 51, 62, 58, 39, 50, 72, 73, 63, 52, 41, 37, 28, 46, 71, 75, 36, 28, 44, 90, 51, 28, 60, 18, 47, 40.

1. Describe or demonstrate a means of displaying the results more clearly.
2. The teacher wants to compare the student's scores with those of another class. Describe a means of organizing the data that would make it easy to compare the two sets of data.

3. The teacher gave grades as follows:

A grade: 90 and above

B grade: 80 to 89

C grade: 70 to 79

D grade: 60 to 69

F grade: 59 and below

Make a table to show how many students achieved each grade

4. Determine if the data is qualitative or quantitative.

- a. The majority of the people in Asia most often wear the color red.
- b. A survey was done among elementary age children to discover their favorite fruit.

5. These are the numbers of cars sold at a local dealer over the last 12 days. Create a Frequency Distribution Table.  
3, 5, 1, 7, 3, 2, 8, 1, 3, 2, 6, 4.

**Solution:**

1. A good start would be to simply organize the numbers in increasing order:

18, 20, 28, 28, 34, 36, 37, 38, 39, 40, 41, 44, 45, 46, 46, 47, 48, 50, 50, 51, 51, 52, 52, 57, 58, 59, 60, 62, 62, 63, 71, 72, 73, 73, 75, 78, 82, 90, 91.

Now we can see at a glance that the numbers range from 18 to 91, with a greater frequency in the mid-range than at the extremes.

2. To compare the scores with another class, it would be convenient to have the number of scores in each range summarized. She might either tally the number of scores between 0 and 10, then 10 and 20, and so on, or just tally the number of A's, B's, etc.

3. The table would look like this:

**TABLE 4.1:**

A	B	C	D	F
2	1	6	4	26

(Either that was a frightfully difficult exam, or the students didn't study well!)

4. These are both qualitative. Neither A, nor B could be expressed as numerical data.

5. To create a frequency distribution table for 3, 5, 1, 4, 3, 2, 2, 1, 3, 2, 5, 4, simply label the values that occur in the set across the top, and the number of occurrences of each in a 2<sup>nd</sup> row beneath, either as numerals or as tally marks:

**TABLE 4.2:**

Value:	1	2	3	4	5
Frequency:	<i>II</i>	<i>III</i>	<i>III</i>	<i>II</i>	<i>II</i>
	2	3	3	2	2

**Practice**

For Q's 1-3, determine if the data is qualitative or quantitative.

1. The average temperature of a particular city is 23 degrees C.
2. Determine if the number of hours a person spends in front of a computer will affect their eye sight.
3. A random survey was done to find out the average speed of cars on a highway.
4. Which letter has the greatest frequency in the following sentence?

THE SUN ALWAYS SETS IN THE WEST.

5. Joe scored the following numbers of goals in their last twenty soccer games: 3, 0, 1, 5, 4, 3, 2, 6, 4, 2, 3, 3, 0, 7, 1, 1, 2, 3, 4, 3.

- a. Organize the values from smallest to greatest
- b. Which number had the greatest frequency?



6. The following number gives the first 31 digits of pi: 3141592653589793238462643383279
- Treating each digit as a separate unit of data, how might you organize the units to prepare for an evaluation of their frequency and range (spread)?
  - How would the units appear after the organization?
  - What is the frequency of the digits 3, 5, and 7?

7. A die was thrown 100 times. The frequency distribution is shown in the following table:

**TABLE 4.3:**

Roll	Frequency
1	21
2	11
3	15
4	19
5	16
6	18

- What is the total frequency of numbers less than 4?
- What percentage of throws of the die were higher than 5?
- How many throws scored greater than 2, but less than or equal to 5?

50 students took a test with a total of 10 possible points. The frequency distribution is shown in the following table:

**TABLE 4.4:**

Score	Frequency
0	1
1	2
2	1
3	3
4	1
5	4
6	9
7	8
8	7
9	10
10	4

- If 60% is a passing score, how many students passed the test?
- How many students scored above 80%?
- What percentage of the students had 5 or more questions correct?
- How many students scored greater than or equal to 4, but less than or equal to 7?

A spinner is in the shape of a regular heptagon marked with the numbers 1 to 7. Sue spun the spinner 50 times and recorded her results:

1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7

12. Create a frequency table with the data.

13. Which spin had the greatest frequency?

14. Which spin had the least frequency?

A teacher decided to survey the students in her class to determine the number of siblings each of them had. The following numbers are the total number of siblings reported by each student in the class: 2, 0, 1, 0, 1, 0, 4, 3, 4, 9, 2, 1, 3, 1, 5, 1, 2, 1, 2, 4, 3, 2, 2, 6, 3, 2, 4, 2, 3, 5

15. Organize the numbers in a manner conducive to the creation of a frequency table.

16. Create a Frequency Table.

17. How many students were surveyed to collect this data?

18. How many families have 4 children or less?

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.1.

## 4.2 Analyzing Data

### Objective

Here you will practice identifying the most appropriate data visualization for a particular set of data and the intended purpose of the study.

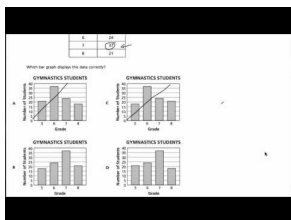
### Concept

How do you decide what sort of graph is most appropriate for a particular application? If you know you have a set of average test scores from a number of different tests and are attempting to compare them, should you use a histogram, box-and-whiskers plot, or a bar chart? If you have 500 yes/no/maybe responses to a survey, should you visualize them with a pie chart, a bar chart, or a frequency polygon?

Choosing an effective data visualization can be a bit daunting; particularly at first, but with practice it will become much less difficult.

### Watch This

This video is a rather detailed discussion of the use of a number of different types of graphs.



### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63582>

<http://youtu.be/c3DBUzSIJR8> ShaunTeaches - Different Types of Graphs

### Guidance

There are quite a number of common types of data visualizations, some of which have been in use for hundreds of years, and using modern technology there are even more. Some of the more modern types are animated and/or interactive, and may continually update themselves as new data is collected via the Internet. There are a number of other lessons in this course that individually detail the creation and interpretation of some of the more common graphing types, but in this lesson we will focus on some of the advantages and disadvantages of each. Having a good idea of the strengths of different methods will help you to choose an appropriate method for your own study.

**TABLE 4.5:**

Graphing Method	Strengths / Advantages	Weaknesses / Disadvantages
-----------------	------------------------	----------------------------

**TABLE 4.5:** (continued)

<b>Histogram</b>	Good for comparing multiple ranges of values, can be visually appealing, good for consolidating large amounts of data into a limited number of categories	Gets cluttered if there are too many categories of data, can take a while to construct - particularly by hand, not ideal for comparing a large number of categories, precise data is not easily displayed since values are grouped
<b>Box-and-Whisker Plot</b>	Quickly demonstrates a five-point summary of the data (mean, quartiles, range), can be used to compare multiple data sets, good for consolidating very large data sets	Individual data usually lost as data is grouped, not often considered as visually appealing as some other graphs, can be somewhat slow to construct
<b>Stem and Leaf Diagram</b>	Maintains accuracy of individual data values, can be easily evaluated to find ranges and notable clusters, good for medium to med-large data sets	Generally not particularly attractive, not convenient for identifying central tendencies for larger data sets, can be confusing for audiences not familiar with the idea.
<b>Scatter Plot</b>	Good for showing increasing/decreasing data trends, maintains accuracy of individual values, makes outliers immediately apparent, faster to construct than some, excellent for bivariate data.	May lead to incorrect generalizations about the data, may take a while to construct if there are many data points, limited flexibility for creating attractive designs, regression should only be applied to continuous variables.
<b>Line Plot</b>	Continuous variables can be represented with a few discrete points, handles bivariate data, moderate flexibility in design, easily interpreted by audiences	Requires continuous data, may not be visually striking
<b>Frequency Polygon</b>	Same as Line Plot above, but with greater visual flexibility for grabbing audience attention	Requires continuous data, anchors on base may be assumed as zero value data points
<b>Pie Chart</b>	Simple to create and interpret, easily visualize relative percentages, great for comparing values from multiple sources, can be very visually striking	Limited uses, requires discrete data, may be overly simplistic

**Example A**

Which of the following types of graphs is the most appropriate for displaying/evaluating the body weights of 300 different pet dogs? Why?

Bar Chart    Scatter Plot    Pie Chart    Histogram

**Solution:** A histogram is the best option here because it allows you to break the data up into as many or as few categories or weight classes as you wish, and clearly shows the comparison between them. Also, a histogram would be a relatively efficient way to track the rather large data set.

**Example B**

Which one of the given graph types is most appropriate for evaluating the relationship between amount of sunlight and plant growth rate?

Bar Chart    Scatter Plot    Pie Chart    Histogram

**Solution:** A scatter plot is the right tool here, since you are comparing two different variables. Since these are both *continuous variables*, you could also evaluate the trend of the data points. If the trend appears to be linear, you could use *linear regression* to identify the average comparison, and then add a linear plot to the graph to illustrate that average.

**Example C**

The relative distribution of several age ranges of trees in a forest: 0-5yrs, 6-15yrs, 16-30yrs, 31yrs and older.

Bar Chart    Scatter Plot    Pie Chart    Histogram

**Solution:** This would be a great use for a pie chart, since you are dealing with the entire population of trees, and are interested in comparing the percentage of the forest represented by each.

**Concept Problem Revisited**

*How do you decide what sort of graph is most appropriate for a particular application? If you know you have a set of average test scores from a number of different tests and are attempting to compare them, should you use a histogram, box-and-whiskers plot, or a bar chart? If you have 500 yes/no/maybe responses to a survey, should you visualize them with a pie chart, a bar chart, or a frequency polygon?*

Choosing the best visual representation of a data set is a matter of identifying the purpose of your study and the type of data you intend to display.

A particular average representing each test could be pretty clearly displayed by a bar chart, and could be made quite striking with appropriate design.

500 results spread over only three categories would be a good use of a pie chart. You have the entire population (members of the survey), and your goal is to show the relative portions of each answer.

**Vocabulary**

**Linear regression** is the process of evaluating a single line best representing a number of individual plots on a scatter plot that shows a linear trend.

**Continuous variables** represent data that could be infinitely divided into smaller and smaller 'pieces'. In other words, data that can take on an infinite number of values in any specified interval. The most common continuous variable is time.

**Guided Practice**

Choose the most appropriate graphing method for each situation, and explain your reasoning:

1. The number of high school diplomas earned in Denver, CO for each year between 1980 and 1990.

Bar Chart    Scatter Plot    Pie Chart    Histogram



2. The proportion of minnows in a pond that are damaged by chemical dumping.

Bar Chart    Scatter Plot    Pie Chart    Histogram

3. The average number of puppies birthed by five breeds of dogs.

Bar Chart    Scatter Plot    Pie Chart    Histogram

4. The number of students in the upper 10%, upper 50%, lower 50%, upper 25%, and lower 25% on finals at a particular university.

Bar Chart    Scatter Plot    Pie Chart    Histogram

5. The number of yellow, red, and white roses found in each of 500 ten by ten foot plots in Central Park, New York.

Bar Chart    Scatter Plot    Pie Chart    Histogram

Solutions:

1. Histogram, this is a comparative study of a limited number of categories of data. Note that although time is continuous, the groups of 1 year each allow us to graph with a histogram.
2. Pie Chart, this is a finite population and you are comparing proportions of a limited number of categories.
3. Bar Chart, since you are dealing with averages of discrete data in a limited number of categories.
4. Bar Chart, your categories overlap (upper 50% includes upper 10% for example), so a pie chart would not be appropriate.
5. Scatter Plot, there are WAY too many categories (individual flower garden plots) to try to use a histogram or bar chart.

## Practice

What type of Graph should you use?

1. Track and compare the different amounts of time you spend playing video games versus doing your homework and practicing piano, over the period of a month.
2. You are trying to prove that your family spends too much money on groceries each month. Your families monthly budget looks like this:
  1. \$1,250 home mortgage
  2. \$500 utilities
  3. \$800 car payments
  4. \$300 entertainment
  5. \$800 groceries
3. You employ 4 sales people. You would like to track their sales over the last quarter.
4. You would like to compare the number of votes that 4 candidates received in the last student council election.
5. You would like to know if there is a relationship between the time you spend studying for a test and the test scores you receive in a class, knowing you study more for classes you enjoy.

6. Choose methods of graphing commonly used to “Compare”. For instance comparing the sales performance of one car to another.
  1. Bar Graphs
  2. Column Graphs
  3. Scatter Plots
  4. Pie Charts
  5. Line Graphs
  6. Data Tables
  7. Box Plots
  8. Histograms
7. Choose types of graphs commonly used to show “Distribution”. For instance, the waiting room times of patients in 5 different doctors’ offices.
  1. Bar Graphs
  2. Column Graphs
  3. Scatter Plots
  4. Pie Charts
  5. Line Graphs
  6. Data Tables
  7. Box Plots
  8. Histograms
8. Choose types of graphs commonly used to show “Parts of a Whole” For instance, number of female viewers of a new gaming website.
  1. Bar Graphs
  2. Column Graphs
  3. Scatter Plots
  4. Pie Charts
  5. Line Graphs
  6. Data Tables
  7. Box Plots
  8. Histograms
9. Choose types of graphs often used to show “Trends over Time”. For instance, number of umbrellas sold over a 365 day period.
  1. Bar Graphs
  2. Column Graphs
  3. Scatter Plots
  4. Pie Charts
  5. Line Graphs
  6. Data Tables
  7. Box Plots
  8. Histograms
10. Choose types of graphs used to show “Deviations”. For instance, sales numbers in an established business when a competitor opens across town.
  1. Bar Graphs
  2. Column Graphs
  3. Scatter Plots
  4. Pie Charts
  5. Line Graphs
  6. Data Tables

7. Box Plots
  8. Histograms
11. Choose types of charts that can be used to show “relationship” For instance, number of tutoring students obtained after report cards are released.
1. Bar Graphs
  2. Column Graphs
  3. Scatter Plots
  4. Pie Charts
  5. Line Graphs
  6. Data Tables
  7. Box Plots
  8. Histogram

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.2.

## 4.3 Relative Frequencies

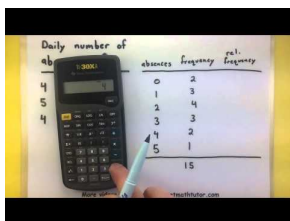
### Objective

Here you will learn about comparing the relative number of times that different data values appear in a set.

### Concept

If you were evaluating a set of data describing the numbers of “A’s”, “B’s”, “C’s”, and “D’s” that students earned on a particular test, and needed to display the data on a *relative frequency table*, how would you go about it?

### Watch This



### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63584>

<http://youtu.be/7jUIt39tUBM> MySecretMathTutor - How to make a relative frequency distribution

### Guidance

Frequency tables are closely related to histograms and stem-and-leaf plots. A relative frequency table is specifically designed to display the ratio of each individual frequency to the total frequency of the data. To begin building a relative frequency table, start by grouping values into categories, classes, or intervals, depending on the type of data. You should try to limit the number of intervals or classes to less than a dozen in most cases, and you can use the square root of the number of actual data points as a guide if you wish.

Once you have all of your data separated into separate classes or categories, (often known as “*binning*”, since the data is divided up into multiple “bins,” one for each specified class, category, or interval), tally the number of values in each category and the total number of values all together.

To calculate the relative frequency of each category, divide the category, class, or interval frequency by the overall frequency. The decimal you get will represent the part of the entire sample that is represented by that category. Once you have calculated all of the relative frequencies for every category, add them up to make sure they total 1.0.

**Note!** If you are graphing the relative frequencies of a *continuous variable*, you will need to specify how to handle any values that fall right on the boundary of a category (also commonly called a class). Here are a couple of ways to do this:

- You can specify on your table that values equal to lower class limits are included in a class, but values equal to upper class limits are not (this is the conventional method). This means that a value of 5 would be considered part of a 5-10 class, but not part of a 1-5 class.

- You can also define each category so that there are no overlapping values:

1-4.99 5-9.99 10-14.99 15-20

### Example A

You are given a bag of marbles in multiple colors, if there are 25 red, 22 yellow, 17 green, and 28 blue marbles, what are the relative frequencies of each color?

**Solution:** Start by totaling the number of marbles:  $25 + 22 + 17 + 28 = 92$  total marbles



Divide the number of each color by the total number of marbles:

$$\frac{25 \text{ red marbles}}{92 \text{ total marbles}} = .272$$

$$\frac{22 \text{ yellow marbles}}{92 \text{ total marbles}} = .239$$

$$\frac{17 \text{ green marbles}}{92 \text{ total marbles}} = .185$$

$$\frac{28 \text{ blue marbles}}{92 \text{ total marbles}} = .304$$

Add your totals together to verify that they equal 1:

$$.272 + .239 + .185 + .304 = 1.0$$

Note that each of the relative frequencies can also be understood as percentages:

$$.272 = 27.2\% \text{ red marbles}$$

$$.239 = 23.9\% \text{ yellow marbles}$$

$$.185 = 18.5\% \text{ green marbles}$$

$$.304 = 30.4\% \text{ blue marbles}$$

$$27.2\% + 23.9\% + 18.5\% + 30.4\% = 100\%$$

### Example B

A police officer is reviewing accident statistics for her city. She notes that there were a total of 23 incidents involving teen drivers between ages sixteen and twenty-one, 19 incidents involving drivers aged twenty-two through twenty-six, 19 involving twenty-seven to forty-year-olds, and 18 for ages above forty-one.

What are the relative frequencies for each age range?





**Solution:** The total number of accidents is:

$$23 + 19 + 19 + 18 = 79 \text{ total accidents}$$

The relative frequencies are:

$$\frac{23 \text{ in age range } 16-21}{79 \text{ total}} = .291$$

$$\frac{19 \text{ in age range } 22-26}{79 \text{ total}} = .241$$

$$\frac{19 \text{ in age range } 27-40}{79 \text{ total}} = .241$$

$$\frac{18 \text{ in age range } 41+}{79 \text{ total}} = .228$$

Verify that the relative frequencies total 1.0:

$$.291 + .241 + .241 + .228 = 1.001 \text{ (due to rounding)}$$

### Example C

A local high school has 150 students who drive to school. Examining the parking lot, you note that there are 25 white cars, 35 red cars, 13 green cars, 19 blue cars, and 58 others.

What are the chances, expressed as percentages that randomly chosen students have each of the different colored cars?

**Solution:** We are given the total number of cars in the question: 150

Divide each of the individual colors by the total and convert the decimal answers to percentages:

$$\frac{25 \text{ white cars}}{150 \text{ total cars}} = .167 = 16.7\%$$

$$\frac{35 \text{ red cars}}{150 \text{ total cars}} = .233 = 23.3\%$$

$$\frac{13 \text{ green cars}}{150 \text{ total cars}} = .087 = 8.7\%$$

$$\frac{19 \text{ blue cars}}{150 \text{ total cars}} = .127 = 12.7\%$$

$$\frac{58 \text{ others}}{150 \text{ total cars}} = .387 = 38.7\%$$

### Concept Problem Revisited

If you were evaluating a set of data describing the numbers of “A’s”, “B’s”, “C’s”, and “D’s” that students earned on a particular test, and needed to display the data on a relative frequency table, how would you go about it?

Add up the number of entries in each category, A, B, C, and D, to get the total number of data points. Divide the number of values in each category by the total to get the relative frequencies. Convert the decimal values to percentages if necessary.

## Vocabulary

A **relative frequency table** compares the number of entries in each of several categories to the number of entries in the entire population.

**Binning** is the common term for the process of dividing data up into multiple categories, classes, or intervals in preparation for graphing.

A **continuous variable** is a variable that can represent *any* value between a given minimum and maximum. Age is a common continuous variable, since age can be measured in infinitely small increments. By contrast, a **discrete variable** can only represent *specific* values in a given range. The number rolled on a standard die is a discrete variable since it can only be one of the numbers 1 - 6, no partials or fractions.

## Guided Practice

- The Sackmore and Headbut village football teams have played each other 50 times. Sackmore has won 10 times, Headbut has won 35 times, and the teams have drawn 5 times. Based on past performance, what is the probability that Sackmore will win the next match?
- Tony estimates that the probability that there will be an empty space in the car park when he arrives at work is  $\frac{4}{5}$ . His estimate is based on 50 observations. On how many of these 50 days was he *unable* to find an empty space in the car park?
- A pair of dice (one red, one green) is cast 30 times, and on 4 of these occasions, the sum of the numbers facing up is 7. What is the relative frequency that the sum is 7?
- The students in a class were asked what kind of music they liked. 18 liked rock, 11 liked pop, 5 liked hip hop, and 8 liked country. Create a frequency and relative frequency table using this information.
- In 1990, there were approximately 10,000 fast food outlets in the US that specialized in Mexican food. Of these, the largest were Taco Bell with 4809 outlets, Taco John's with 430 outlets and Del Taco with 275 outlets. The relative frequency that a fast food outlet that specializes in Mexican food is none of the above is:

Solutions:

- So far, Sackmore has won 35 out of the 50 matches. We can write this as a fraction, which (reduced) is:  $\frac{7}{10}$ . This fraction isn't really the probability of Sackmore winning, but it is an *estimate* of that probability. We say that the *relative frequency* of Sackmore winning is  $\frac{7}{10}$ .
- If Tony has figured that he *is* able to find a space 4 of every 5 times he arrives, then he *is not* able to find a space 1 in every 5 times. If we set the ratio:  $\frac{1}{5} = \frac{x}{50}$ , we can solve for  $x$  to find that he did not have a space 10 times.
- Out of thirty throws, four of them were 7's. The relative frequency is  $\frac{4}{30}$  or  $\frac{2}{15}$ .
- To create the frequency table, we just need one column for each category:

**TABLE 4.6:**

Rock	Pop	Hip Hop	Country
18	11	5	8

To convert to a relative frequency table, just divide each frequency by the total:

**TABLE 4.7:**

Rock	Pop	Hip Hop	Country
$\frac{18}{42} = .43$	$\frac{11}{42} = .26$	$\frac{5}{42} = .12$	$\frac{8}{42} = .19$

- The likelihood that a restaurant is *not* one of the top three would equal the number of Mexican fast food restaurants

that are not one of the three:  $10,000 - 4809 - 430 - 275 = 4486$ , divided by the total number of Mexican fast food restaurants, **10,000**:

$$\frac{4,486}{10,000} = .4486 \text{ or } 44.86\%$$

### Practice

30 Students in a class surveyed each other to find out their favorite movie series, and recorded the results in a table like the one shown below.

**TABLE 4.8:**

Movie Series	Number of Likes
Twilight	7
Lord of the Rings	5
Pirates of the Caribbean	9
Harry Potter	6
Narnia	2
High School Musical	1

1. What was the relative frequency for Narnia?
2. What was the relative frequency for Pirates of the Caribbean?
3. 100 people were asked whether they were left-handed. 8 people answered yes. What is the relative frequency of left-handed people in the survey?
4. The relative frequency of getting a white candy from a particular bag is 0.3. If the bag contains 100 candies, estimate the number of whites.
5. Kyle observed 80 cars as they drove by his bedroom window. 24 of them were red. What is the relative frequency of red cars?
6. The relative frequency of rain in April is .6. There are 30 days in April. Estimate the number of days of rain expected in April.

Use the table below listing the heights of 100 male semiprofessional soccer players.

**TABLE 4.9:**

HEIGHTS (INCHES)	FREQUENCY OF STUDENTS	RELATIVE FREQUENCY
59.95-61.95	5	
61.95-63.95	3	$\frac{3}{100} = 0.03$
63.95-65.95		$\frac{15}{100} = 0.15$
65.95-67.95	40	$\frac{40}{100} = 0.40$
67.95-69.95	17	
69.95-71.95	12	$\frac{12}{100} = 0.12$
71.95-73.95		$\frac{7}{100} = 0.07$
73.95-75.95	1	$\frac{1}{100} = 0.01$
	Total = 100	Total =

7. Fill in the blanks and check your answers.

8. The percentage of heights that are from 67.95 to 71.95 inches is:
9. The percentage of heights that are from 67.95 to 73.95 inches is:
10. The percentage of heights that are more than 65.95 inches is:
11. The number of players in the sample who are between 61.95 and 71.95 inches tall is:
12. What kind of data does this chart highlight, qualitative or quantitative?
13. What is the height interval for the players who fall under the frequency of .03?

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.3.

## 4.4 Cumulative Frequencies

### Objective

Here you will learn about organizing and graphing the running totals of multiple categorical frequencies.

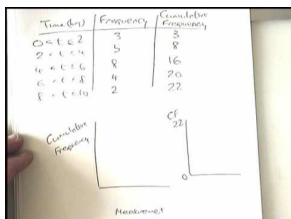
### Concept

A school lunch break is divided into multiple sessions, the 1<sup>st</sup> group, consisting of 20 girls and 50 boys, begins lunch at 11:30 am, the 2<sup>nd</sup> group, with 35 girls and 30 boys begins at 11:40, and the 3<sup>rd</sup> group, 40 girls and 25 boys, at 11:50.

Assuming that none of the students from an early group leave the cafeteria before everyone goes back to class at 12:10, what is the relative frequency of girls in the cafeteria at 11:35? What about at 11:45? 11:55?

In this lesson we discuss *cumulative frequencies*, and by the end of the lesson, this sort of question will be no problem at all.

### Watch This



### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63600>

<http://youtu.be/D1s0XPMnHzo> Maths520 - Cumulative Frequency Part 1

### Guidance

Cumulative frequencies describe a sort of 'running total' of frequencies in a frequency distribution. You will find cumulative frequencies in many real-world situations, since we often need to collect data for a larger study by conducting several smaller studies.

To calculate a cumulative frequency, simply create a frequency distribution table, then add the frequency from the first category to that of the second category, then add that total frequency to the third category, and so on. Because cumulative distributions don't generally show different combinations of categories, it is a good idea to arrange your categories in a suitable order so that the cumulative frequencies are most likely to be of use as the running total increases.

A *relative cumulative frequency* table shows how the cumulative frequency after each successive interval compares to the *total* frequency. To create a relative cumulative frequency table, calculate the relative frequency of each interval or category, and then add the relative frequency of each category to all the prior ones.

### Example A



Create a *cumulative frequency table* showing the number of hours per week that Brian watches television, based on the given information.

Brian's T.V. Time



- Monday: 1.5 hrs
- Tuesday: 2.25 hrs
- Wednesday: 2 hrs
- Thursday: 1.75 hrs
- Friday: 1.5 hrs
- Saturday: 3.25 hrs
- Sunday: 2.5 hrs

Solution:

To find the cumulative frequency of hours of T.V. watched, add the frequency of each day to the total frequency of the day before:

**TABLE 4.10:**

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Cumulative Frequency	1.5 hrs	$1.5 + 2.25 = 3.75$ hrs	$3.75 + 2 = 5.75$ hrs	$5.75 + 1.75 = 7.5$ hrs	$7.5 + 1.5 = 9$ hrs	$9 + 3.25 = 12.25$ hrs	$12.25 + 2.5 = 15$ hrs

### Example B

A car dealer is calculating the total sales for the past month, and wants to identify the percentage of monthly sales that have occurred after weeks 1, 2, 3, and 4. Create a *relative cumulative frequency* table to show the information the dealer wants.

**TABLE 4.11:**

Week Number	Cars Sold
1	21
2	12
3	17
4	24

**Solution:** First total up the sales for the entire month:

$$21 + 12 + 17 + 24 = 74 \text{ cars}$$

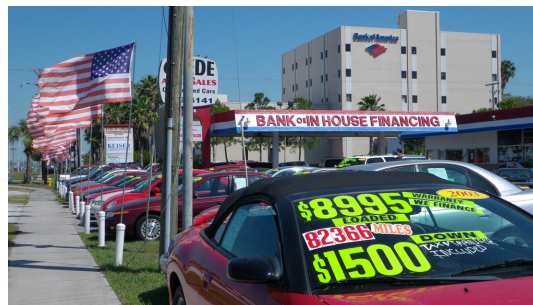
Then find the relative frequencies for each week by dividing the number of cars sold that week by the total:

- The relative frequency for the first week is:  $\frac{21}{74} = .284$
- The relative frequency for the second week is:  $\frac{12}{74} = .162$
- The relative frequency for the third week is:  $\frac{17}{74} = .23$
- The relative frequency for the fourth week is:  $\frac{24}{74} = .324$

To find the relative cumulative frequencies, start with the frequency for week 1 and for each successive week, total all of the previous frequencies:

**TABLE 4.12:**

Week Number	Cars Sold	Relative Frequency	Cumulative Frequency
1	21	.284	.284
2	12	.162	.284 + .162 = .446
3	17	.23	.446 + .23 = .676
4	24	.324	.676 + .324 = 1.0



Note that the first relative cumulative frequency is always the same as the first relative frequency, and the last relative cumulative frequency is always equal to 1.

**Example C**

Create a cumulative and relative cumulative frequency table for the following data:

Number of customers in store on Black Friday each year for the years 1995 - 2005:

47, 49, 48, 54, 57, 52, 61, 65, 67, 66, 70

**Solution:** Label a table with the years 1995 - 2005 across the top and frequency, cumulative frequency, and relative cumulative frequency down the side. Leave 3 blanks underneath each year, and input the frequency for each year:

**TABLE 4.13:**

Year:	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Frequency	47	49	48	54	57	52	61	65	67	66	70
Cumulative Frequency											
Relative Cumulative Frequency											

To calculate the cumulative frequencies, sum the total of all previous frequencies under the frequency for each year:

**TABLE 4.14:**

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
------	------	------	------	------	------	------	------	------	------	------	------

TABLE 4.14: (continued)

<b>Frequency</b>	47	49	48	54	57	52	61	65	67	66	70
<b>Cumulative Frequency</b>	47	96	144	198	255	307	368	433	500	566	636
<b>Relative Cumulative Frequency</b>											

To calculate the relative cumulative frequencies, divide the cumulative frequency for each year by the total cumulative frequency (636 customers, the same as the value for the last year, 2005).

TABLE 4.15:

<b>Year</b>	<b>1995</b>	<b>1996</b>	<b>1997</b>	<b>1998</b>	<b>1999</b>	<b>2000</b>	<b>2001</b>	<b>2002</b>	<b>2003</b>	<b>2004</b>	<b>2005</b>
<b>Frequency</b>	47	49	48	54	57	52	61	65	67	66	70
<b>Cumulative Frequency</b>	47	96	144	198	255	307	368	433	500	566	636
<b>Relative Cumulative Frequency</b>	.074	.151	.226	.311	.401	.483	.579	.681	.786	.890	1.0

### Concept Problem Revisited

A school lunch break is divided into multiple sessions, the 1<sup>st</sup> group, consisting of 20 girls and 50 boys, begins lunch at 11:30 am, the 2<sup>nd</sup> group, with 35 girls and 30 boys begins at 11:40, and the 3<sup>rd</sup> group, 40 girls and 25 boys, at 11:50.

Assuming that none of the students from an early group leave the cafeteria before everyone goes back to class at 12:10, what is the relative frequency of girls in the cafeteria at 11:35? What about at 11:45? 11:55?

First create a table showing each session and the number of boys and girls in each session:

TABLE 4.16:

Session 1 - 11:30		Session 2 - 11:40		Session 3 - 11:50	
Boys	Girls	Boys	Girls	Boys	Girls
50	20	30	35	25	40

Next, add a row to the bottom of your table for cumulative relative frequency of girls:

TABLE 4.17:

Session 1		Session 2		Session 3	
Boys	Girls	Boys	Girls	Boys	Girls
50	20	30	35	25	40
Girls: $\frac{20}{70} = .286$ or 28.6 %		Girls: $\frac{55}{135} =$ .407 or 40.7%		Girls: $\frac{95}{200} =$ .475 or 47.5%	

## Vocabulary

**Relative frequency** compares the frequency of a particular value to the entire sample.

**Relative cumulative frequency** is a running total of *relative* frequencies of all values up to and including the current category.

**Cumulative frequency** is a running total of all *non-relative* frequencies up to and including the current category (*not* as compared to the total).

## Guided Practice

For 25 days, the snow depth at Whiteout Mountain was measured (to the nearest inch) and recorded as follows:

242, 228, 217, 253, 239, 266, 242, 251, 240, 223, 219, 246, 260, 258, 225, 234, 230, 249, 245, 254, 243, 235, 231, 257

1. Determine a reasonable class interval
2. Set up a frequency distribution table
3. Find the frequency for each class interval
4. Calculate the cumulative frequencies progressively from left to right
5. Use the information gathered from the frequency distribution table to plot a cumulative frequency graph.

Solutions:

1. The values range 217 - 266, a class interval of 10 would break the data up into a reasonable 6 bins.
2. and 3: Create a table with the intervals noted across the top, and the frequency of each below:

**TABLE 4.18:**

210-220	220-230	230-240	240-250	250-260	260-270
2	3	5	7	5	2

4.

**TABLE 4.19:**

Interval:	210-220	220-230	230-240	240-250	250-260	260-270
<b>Frequency:</b>	2	3	5	7	5	2
<b>Cumulative Frequency:</b>	$\frac{2}{24}$ 8.33%	$\frac{5}{24}$ 20.83%	$\frac{10}{24}$ 41.66%	$\frac{17}{24}$ 70.83%	$\frac{22}{24}$ 91.66%	$\frac{24}{24}$ 100%

## Practice Questions:

1. A weather forecaster highlights the lows over-night for the past 30 days in a small town in Wisconsin. The temperature readings are given in Fahrenheit, and are shown below. Use the data to complete the frequency table.

41°, 58°, 61°, 54°, 49°, 46°, 52°, 58°, 67°, 43°, 47°, 60°, 52°, 58°, 48°,  
44°, 59°, 66°, 62°, 55°, 44°, 49°, 62°, 61°, 59°, 54°, 57°, 58°, 63°, 60°

**TABLE 4.20:**

Interval	Tally	Frequency
----------	-------	-----------

**TABLE 4.20:** (continued)

40 - 44		
45 - 49		
50 - 54		
55 - 59		
60 - 64		
65 - 69		

2. The following represents scores that a class received on their most recent Biology test. Complete the frequency table below.

58, 79, 81, 99, 68, 92, 76, 84, 53, 57, 81, 91, 77, 50, 65, 57, 51, 72, 84, 89

**TABLE 4.21:**

Interval	Tally	Frequency
50 - 59		
60 - 69		
70 - 79		
80 - 89		
90 - 99		

3. James received the following scores on his quizzes in US History over the course of 1 year. Complete the frequency table below using the scores:

85, 72, 97, 81, 77, 93, 100, 75, 86, 70, 96, and 80.

**TABLE 4.22:**

Interval/Grades	Tally	Frequency
61 - 70		
71 - 80		
81 - 90		
91 - 100		

4. Sue competed against 14 others in a time trial for the 400-meter run at the state finals. Their times have been recorded in the table below. What percent of the runners completed the time trial between 50 and 53.9 seconds?

**TABLE 4.23:** 400 Meter Run Time Trails

Interval	Tally	Frequency
50.0 - 50.9		0
51.0 - 51.9		2
52.0 - 52.9	/	6
53.0 - 53.9		3
54.0 - 54.9		4

5. The following data is the starting weights of 30 adults, in pounds, who participated in a study on weight loss. Use the data to create a cumulative frequency table. Determine appropriate intervals for the weights given.

195, 206, 100, 98, 150, 210, 195, 106, 195, 168, 180, 212, 104, 195, 100, 216, 195, 209, 112, 99, 206, 116, 195,

100, 142, 100, 135, 98, 160, 155

**TABLE 4.24:** Weight of Adults before diet program participation

Interval	Frequency	Cumulative Frequency

6. The following table shows the weights in pounds for students attending a “Get Fit” summer program. Use the data to find cumulative frequency.

**TABLE 4.25:** Weights of “Get Fit” camp students

Interval	Frequency	Cumulative Frequency
91-100	6	
101-110	3	
111-120	0	
121-130	3	
131-140	0	
141-150	2	
151-160	2	

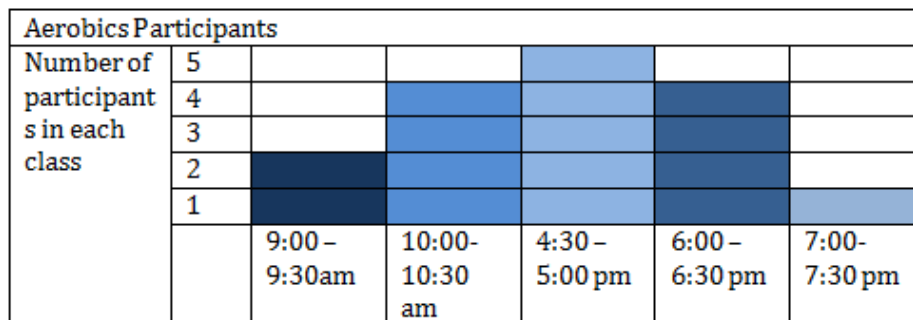
7. 20 students were asked how many meals their family eats out in one week. The results are listed below. Complete the frequency table using this data.

{6, 5, 4, 5, 0, 7, 1, 5, 4, 4, 3, 2, 2, 3, 2, 4, 3, 4, 0, 7}

**TABLE 4.26:** Number of Meals Out

Interval	Tally	Frequency	Cumulative Frequency
0-1			
2-3			
4-5			
6-7			

8. The following graph shows the participants in Jane’s aerobics classes. How many total students does she teach on the given day?



9. The following table shows a cumulative frequency distribution of the ages of tri-athletes. According to the table, how many tri-athletes are in their forties?

**TABLE 4.27:** Cumulative Frequency Distribution of Tri-Athlete Ages

Age Group	Total
20 - 29	8
20 - 39	18
20 - 49	25
20 - 59	31
20 - 69	35

10. The quiz scores for a job placement math assessment for the 10 applicants were: 61, 67, 81, 83, 87, 88, 89, 90, 98, and 100. Which frequency table is accurate for this set of data?

**TABLE 4.28:**

Interval	Freq	Interval	Freq	Interval	Freq	Interval	Freq
61 - 70	2	61 - 70	2	61 - 70	2	61 - 70	2
71 - 80	0	71 - 80	2	71 - 80	0	71 - 80	2
81 - 90	8	81 - 90	8	81 - 90	6	81 - 90	7
91 - 100	10	91 - 100	10	91 - 100	2	91 - 100	10

It is that time of year again. It is Easter, and it is time to hunt eggs. The kids have been divided up into age groups, and their eggs hidden in different areas of a large park. The adult leaders have gathered the data from the three age groups: “Kinders”, “Pee Wee”, and “Big Eggs”, and compiled the results into frequency tables. The problem is, each of the adult leaders reported their information differently.

For questions 11-13, use the data gathered by each adult leader and follow the instructions for interpreting the data given by each.

11. Team “Kinders” coach reported the following results for the number of eggs found by the kiddos in their group, convert them to a cumulative frequency chart.

**TABLE 4.29:**

Time (s) in minutes	Frequency of Eggs Found
0 - 5	20
5 - 10	35
10 - 15	15
15 - 20	22
20 - 25	8
25 - 30	9

12. Team “Pee Wee” also did a great job of finding eggs. Their results were recorded in a cumulative frequency table, convert to a frequency table.

**TABLE 4.30:**

Time (s) in minutes	Cumulative Frequency of Eggs Found
0 - 5	20
5 - 10	27

**TABLE 4.30:** (continued)

10 - 15	31
15 - 20	31
20 - 25	37
25 - 30	50

13. The “Big Eggs” were the champions the last three years running. They are known for always collecting the same amount of eggs in each interval. This year they collected a total of 120 eggs.

- a. Complete a frequency chart
- b. Complete a cumulative frequency chart

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.4.



## 4.5 Creating Histograms

### Objective

Here you will learn how to create *histograms* from data.

### Concept

Consider the data from a previous lesson regarding the number of cars sold during each week of a particular month:

**TABLE 4.31:**

Week Number	Cars Sold
1	21
2	12
3	17
4	24

How would you go about displaying this data as a histogram?

By the end of this lesson, I think you'll see that converting properly organized data into a visual format is really quite straightforward.

### Watch This



#### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63626>

[http://youtu.be/KCH\\_ZDygrm4](http://youtu.be/KCH_ZDygrm4) MySecretMathTutor - How to Make a Histogram

### Guidance

A histogram is really just a rather specific form of a bar chart where the bars are designed to be proportional in area to the frequency and in width to the *interval* that is represented by the bar. The most immediately obvious difference is that there is also no gap between bars, allowing histograms to display continuous data, if it has been organized into intervals.

Creating a histogram is relatively easy once your data has been properly organized, so clearly laying out your data is an important first step. Organizing the data into a table before actually constructing your histogram may also help.

You will want to be sure that you have clearly separated your data into intervals or categories. The number of divisions may depend on your particular data set, but should be few enough not to be confusing to your audience. A

common rule of thumb is to aim for a number of intervals approximately equal to the square root of your number of observations or data points, and generally should be between 5 and 10 intervals. The width of each interval should be approximately the range of your data divided by the number of intervals. In other words, if you have 25 data points ranging from values of 1 to 100, you would expect to have perhaps 5 intervals of 20 units in width each.

### Example A

Create a histogram from the following data:

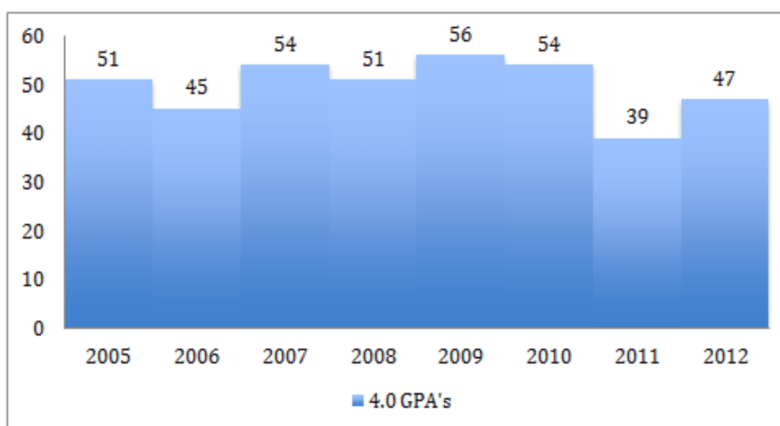
**TABLE 4.32:**

Year	Number of 4.0 G.P.A.'s
2005	51
2006	45
2007	54
2008	51
2009	56
2010	54
2011	39
2012	47

**Solution:** The data in this table is already conveniently arranged for display as a histogram. Note that there are only 8 intervals of 1 year each and a single frequency count for each interval. The only thing we need to identify before we actually place this data into a histogram is the frequency range.

The lowest frequency on the table is 39, in year 2011. The highest frequency is 56, in year 2009. That means we have a vertical range of  $56 - 39 = 17$  units. Let's plan to label the frequencies from 0 to 60 up the left side of our histogram.

To convert this table into a histogram, all we need to do is create a chart with the intervals (the years 2005 through 2012) across the bottom along the  $x$ -axis, and the frequencies (the number of 4.0 G.P.A.'s each year) up the side.



### Example B

Create a histogram demonstrating the number of prank calls reported over time using the data below.

**TABLE 4.33:**

Year	Number of Pranks	Year	Number of Pranks
1970	24	1985	24
1971	42	1986	28

TABLE 4.33: (continued)

1972	38	1987	38
1973	42	1988	40
1974	42	1989	42
1975	27	1990	26
1976	25	1991	39
1977	37	1992	32
1978	27	1993	26
1979	33	1994	31
1980	25	1995	36
1981	32	1996	37
1982	40	1997	38
1983	38	1998	28
1984	26	1999	33

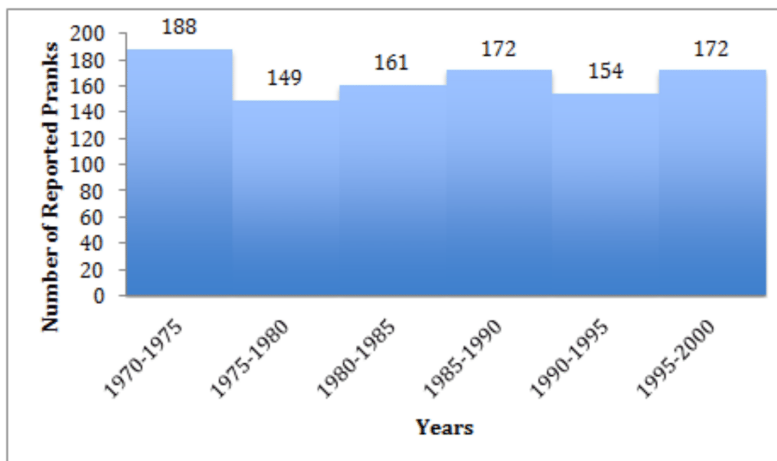
**Solution:** This table has much more data than Example A. If we were to follow the same procedure, we would end up with 30 intervals, which would make a rather crowded and confusing histogram. We can solve this by *binning* the data before we import it into our histogram.

*Binning* data means to group the given intervals or categories into broader ranges (bins) in order to limit the number of different intervals. In this case, we have a total of 30 data points, so our 'square-root rule of thumb' would suggest 5-6 bins. If we opt for 5 bins, we would have  $\frac{30}{5} = 6$  years per bin, and if we opt for 6 bins, we would have  $\frac{30}{6} = 5$  years per bin. 5-year intervals are common, and will be easy for your audience to understand. If we group our original data into 5-year intervals, it looks like this:

TABLE 4.34:

Interval	Number of Pranks
1970 - 1975	$24 + 42 + 38 + 42 + 42 = 188$
1975 - 1980	$27 + 25 + 37 + 27 + 33 = 149$
1980 - 1985	$25 + 32 + 40 + 38 + 26 = 161$
1985 - 1990	$24 + 28 + 38 + 40 + 42 = 172$
1990 - 1995	$26 + 39 + 32 + 26 + 31 = 154$
1995 - 2000	$36 + 37 + 38 + 28 + 33 = 172$

Now we can take our binned data and display it as a histogram with our 5-year intervals along the bottom and number of reported pranks (the frequencies) up the side:



### Example C

Create a histogram to display the number of shoe types in several price ranges, using data from the table below:

**TABLE 4.35:**

Shoe Type	Price	Shoe Type	Type
sandals	\$25	loafers	\$44
sneakers	\$63	high heels	\$28
sneakers	\$43	boots	\$26
sneakers	\$38	high heels	\$43
loafers	\$62	sandals	\$62
sneakers	\$45	loafers	\$30
sneakers	\$56	high heels	\$51
sneakers	\$25	sneakers	\$58
loafers	\$44	high heels	\$63
boots	\$30	loafers	\$43
high heels	\$36	loafers	\$31
loafers	\$29	sneakers	\$56
loafers	\$62	boots	\$54
loafers	\$48	high heels	\$55
sneakers	\$45	sandals	\$40

**Solution:** The first step is to identify your categories or intervals. In this situation, we are expected to compare the frequency of shoe types in price range intervals, so we will need to decide what our intervals should be by dividing up the price ranges appropriately.

The lowest price in the data set is \$25, and the highest is \$63. This gives us a *range* of \$38, so we can round that up to \$40. \$40 divides conveniently by 5, so we could go with 8 intervals of \$5 each. That may be a bit more detailed than we really need since there are only 30 data points, which would suggest 5 or 6 intervals rather than 8, but we are still under 10, and the \$5 intervals will be easy for our audience to identify with.



Next we need to identify a range of frequencies for our selected intervals, which means we will need to count the number of entries in each interval using our question data. At this point we can either 'hunt through' the data, counting as we go, or we can re-organize the original data by price to simplify the counting process. Since we are not actually using the data regarding the names of each shoe type, only the frequency based on price, we can just organize the prices in dollars by ascending order:

25, 25, 26, 28, 29, 30, 30, 31, 36, 38, 40, 43, 43, 43, 44, 44, 45, 45, 48, 51, 54, 55, 56, 56, 58, 62, 62, 62, 63, 63

Then we can group the list according to our chosen interval of \$5:

25, 25, 26, 28, 29 | 30, 30, 31 | 36, 38 | 40, 43, 43, 43, 44, 44 | 45, 45, 48 | 51, 54 | 55, 56, 56, 58 | 62, 62, 62, 63, 63

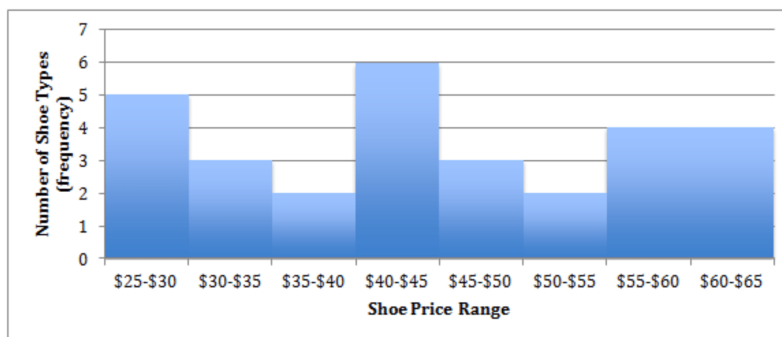
By convention, a value equal to the lower limit of an interval is included, but a value equal to the upper limit is not (upper and lower interval limits are also called *class limits*). For example, a \$40 entry is included in the \$40-\$45 interval since \$40 is the lower limit of that interval, but the two \$45 entries are included in the next interval up (\$45-\$50), since \$45 would be the upper limit of the \$40-\$45 interval.

Finally, we take the grouped data and summarize by frequency in each interval:

**TABLE 4.36:**

Interval (Price range)	Frequency (Number of shoe types)
\$25-\$30	5
\$30-\$35	3
\$35-\$40	2
\$40-\$45	6
\$45-\$50	3
\$50-\$55	2
\$55-\$60	4
\$60-\$65	4

Finally we construct our histogram frame with intervals along the bottom and frequencies up the side:



### Concept Problem Revisited

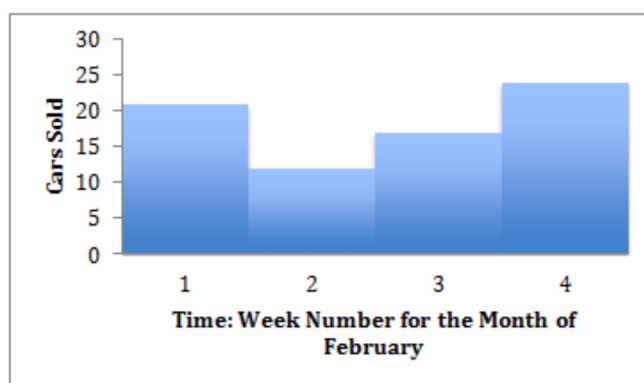
Consider the data from a previous lesson regarding the number of cars sold during each week of a particular February:

**TABLE 4.37:**

Week	Cars
1	21
2	12
3	17
4	24

How would you go about displaying this data as a histogram?

Compared to the examples you completed during the lesson, this is a piece of cake! The data is already organized into 4 convenient bins, and the number of cars sold has been totaled for each bin. Just create your histogram with weeks across the bottom and number of cars sold up the side:



### Vocabulary

A **Histogram** is a specific form of a bar chart with zero space between each bar, a bar area that is proportional to the frequency it represents, and a bar width equal to the data interval.

An **Interval** is a range of data. Grouping data into intervals can be beneficial in a number of ways, including simplifying the appearance and minimizing the effect of individual measurement errors.

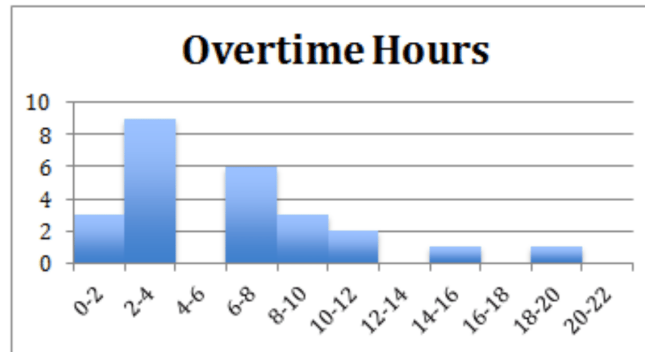
A **Range** is a value representing the difference between the least and greatest value in a data set. The range can be found by subtracting the smallest value from the largest.

**Binning** is the process of grouping data ranges into appropriate intervals. There is no all-around best number of bins, and different numbers of bins can reveal different things about a data set.

**Class limits** are, collectively, the upper and lower limit of an interval. A **class mark** is the middle value, or average of the class limits.

### Guided Practice

The head of human resources wants to know how many of the department heads work overtime in a given month. The data is displayed in the following histogram. Use it to answer questions 1 - 3.



1. What do the numbers on the horizontal axis represent?
2. What do the numbers on the vertical axis represent?
3. What percent of department heads work 6 or more hours of overtime in a given month?
4. Here are the driving scores of 15 new drivers. Create a histogram from their scores:  
88, 48, 60, 51, 57, 85, 69, 75, 97, 72, 71, 79, 65, 63, 73
5. Using the data from the created histogram, identify the percentage of people who did not pass the driving test if the minimum score was 70%.

Solutions:

1. The horizontal axis shows intervals of hours worked overtime.
2. The vertical axis displays the number of department heads falling within each interval.
3. 52%. The histogram displays information about 25 employees. Out of these 25 employees 13 worked more than 6 hours overtime. We divide  $\frac{13}{25}$  and come up with .52, or 52% who work more than 6 hours of overtime.
4. First we should determine how to break the range of values into intervals. In this instance, since our data set consists of driving scores, it would make sense to choose intervals of 10 points: 40-50, 50-60, ... 90-100, since most tests are determined by a certain percentage. By counting how many of the 25 observations fall in each of the intervals, we get the following table:

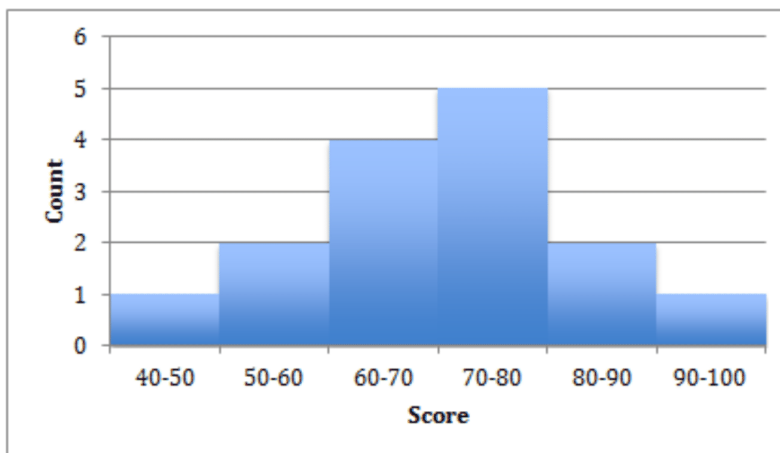
**TABLE 4.38:** Driving Scores

Score	Count
[40-50)	1
[50-60)	2
[60-70)	4
[70-80)	5

**TABLE 4.38:** (continued)

[80-90)	2
[90-100]	1

To construct the histogram from this table, we plot the intervals on the  $X$ -axis, and represent the number of observations in each interval (the frequency of the interval) on the  $Y$ -axis, by the height of a rectangle located above the interval:



5. 47% - First we must find the total number of people who took the driving test. It looks like this:  $1 + 2 + 4 + 5 + 2 + 1 = 15$ . 15 people took the test. Now we identify how many received below 70 %,  $1 + 2 + 4 = 7$ . To find out the percentage, we divide the number of people who earned less than 70%, or 7, by the total number of people who took the driver's test, 15.  $\frac{7}{15} = .47$ . We can say that 47% of the people who took the driver's test did not pass.

### Practice

The following numbers need to be used in creating a histogram. Use this data to answer questions 1-4

9.3, 10.6, 10.6, 10.6, 10.2, 10.7, 9.9, 10.5, 10.4, 8.8, 9.6, 9.8, 9.2, 10.9, 10.0, 11.4, 10.8, 8.9, 10.4, 8.9, 9.6, 8.4, 11.2, 10.3, 10.1, 9.6, 11.1, 9.4, 9.4, 10.6, 8.9, 10.8, 9.2.

- Identify the range for the above data.
- Based on the data, what would be an appropriate number of Intervals?
  - 7-10
  - 11-15
  - 13-20
- If the width of an interval is .4, and the lowest reading is 8.4, it is the starting point. What will be the upper end of the first interval? The second, third and so on?
- Create a histogram for the data set above. Remember that numbers that fall on the high end of one interval and the low end of the next interval should be placed in the lower interval. For instance in the interval 8.4 - 8.8, only 8.4, 8.5, 8.6, and 8.7 belong, 8.8 would be the lowest value in the next interval.
- The following table shows a cumulative frequency distribution of the ages of tri-athletes. Convert the data to standard frequencies, and create a histogram to graph the absolute frequencies.



**TABLE 4.39:** Cumulative Frequency Distribution of Tri-Athlete Ages

Age Group	Total
20-29	8
20-39	18
20-49	25
20-59	31
20-69	35

6. The quiz scores for a job placement math assessment for the 10 applicants were: 61, 67, 81, 83, 87, 88, 89, 90, 98, and 100. Using the table below create a histogram to graph the data.

**TABLE 4.40:**

Interval	Frequency
61-70	2
71-80	0
81-90	6
91-100	2

7. Create a Histogram using the following list of overnight lows:

41°, 58°, 61°, 54°, 49°, 46°, 52°, 58°, 67°, 43°, 47°, 60°, 52°, 58°, 48°, 44°, 59°, 66°, 62°, 55°, 44°, 49°, 62°, 61°, 59°, 54°, 57°, 58°, 63°, 60°

And the following intervals:

40-44

45-49

50-54

55-59

60-64

65-69

8. The following represents scores that a class received on their most recent Biology test. Create a histogram from the scores.

58, 79, 81, 99, 68, 92, 76, 84, 53, 57, 81, 91, 77, 50, 65, 57, 51, 72, 84, and 89.

Use 10-point intervals starting with the interval 50-59.

9. James received the following scores on his quizzes in US History over the course of 1 year. Create a histogram from the scores.

85, 72, 97, 81, 77, 93, 100, 75, 86, 70, 96, and 80.

10. Sue competed in a time trial for the 400-meter run at the state finals. She ran against 14 others. Their times have been recorded in the table below. Complete the chart and create a histogram to represent the data.

**TABLE 4.41:** 400 Meter Run Time Trials

Interval	Tally	Frequency
50.0-50.9		
51.0-51.9		

**TABLE 4.41:** (continued)

52.0-52.9		
53.0-53.9		
54.0-54.9		

11. The following data is the weight of 30 adults, in pounds, who participated in a study on weight loss. Use the data to create a cumulative frequency table. Determine appropriate intervals for the weights given.

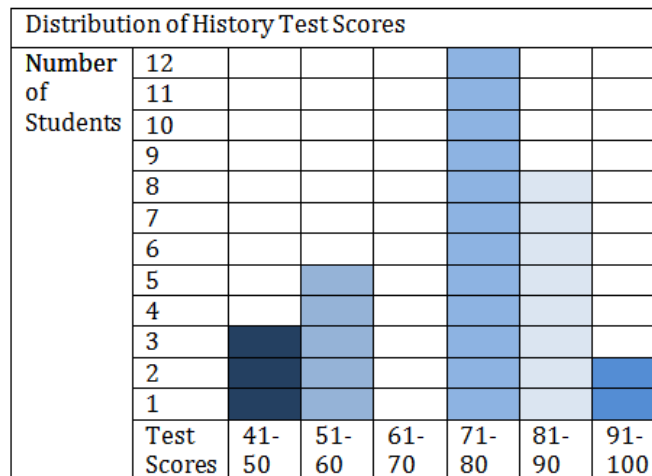
195, 206, 100, 98, 150, 210, 195, 106, 195, 168, 180, 212, 104, 195, 100, 216, 195, 209, 112, 99, 206, 116, 195, 100, 142, 100, 135, 98, 160, 155

12. The following table shows the weights in pounds for students attending a “Get Fit” summer program. Create a histogram to graph the data.

**TABLE 4.42:**

Interval	Frequency
91-100	6
101-110	3
111-120	0
121-130	3
131-140	0
141-150	2
151-160	2

13. The graph below shows the distribution of scores of 30 students on a history exam. Complete the frequency table below it using the data.



**TABLE 4.43:**

Test Scores	Frequency
91-100	
81-90	
71-80	
61-70	
51-60	

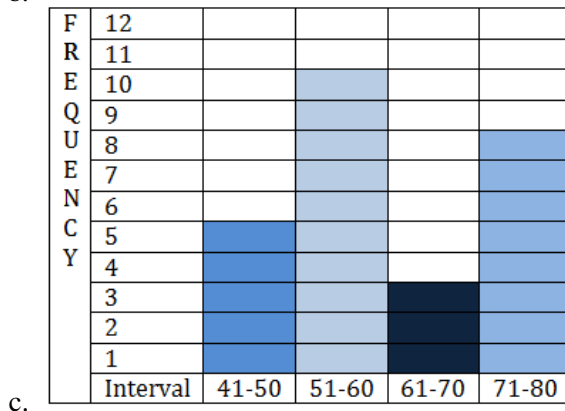
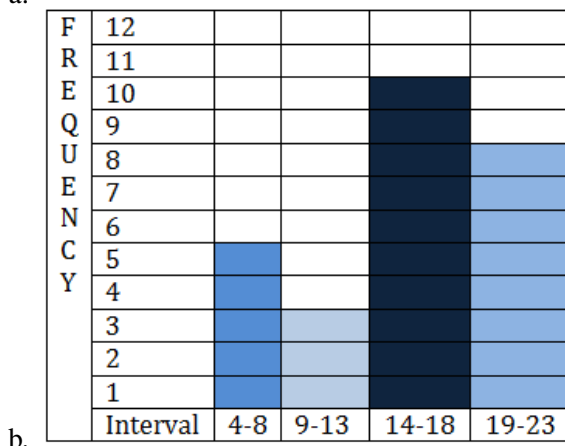
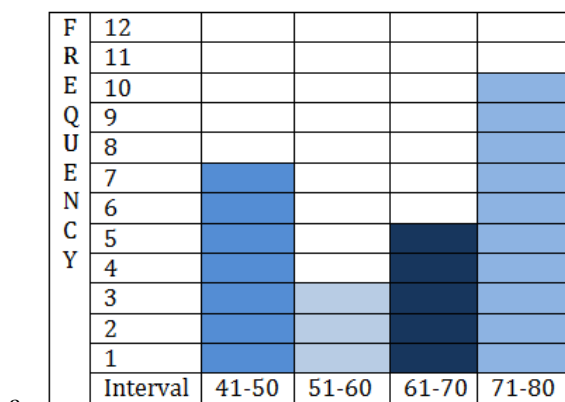
**TABLE 4.43:** (continued)

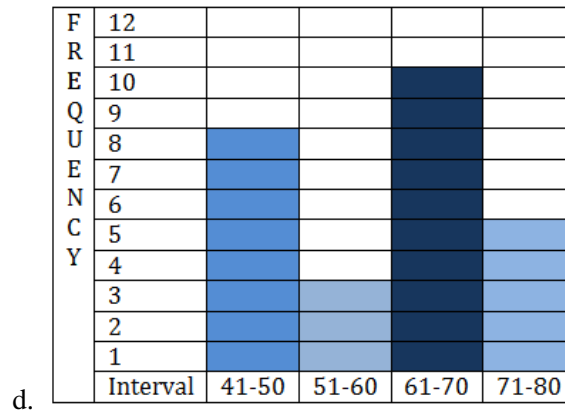
41-50	
-------	--

14. Which one of the following histograms represents the data in the table below?

**TABLE 4.44:**

Interval	Frequency
4-8	8
9-13	3
14-18	10
19-23	5





A teacher was asked how many students they had to ask to stop texting on their cell phones during each of their class periods on a given day.

The following data was collected.

**TABLE 4.45:**

Class Period	Frequency
1	16
2	11
3	5
4	3
5	1

15. Convert the data to a histogram.

16. How many class periods did a teacher have to ask less than 9 times?

17. How many times total did the teacher have to ask students to stop texting?

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.5.

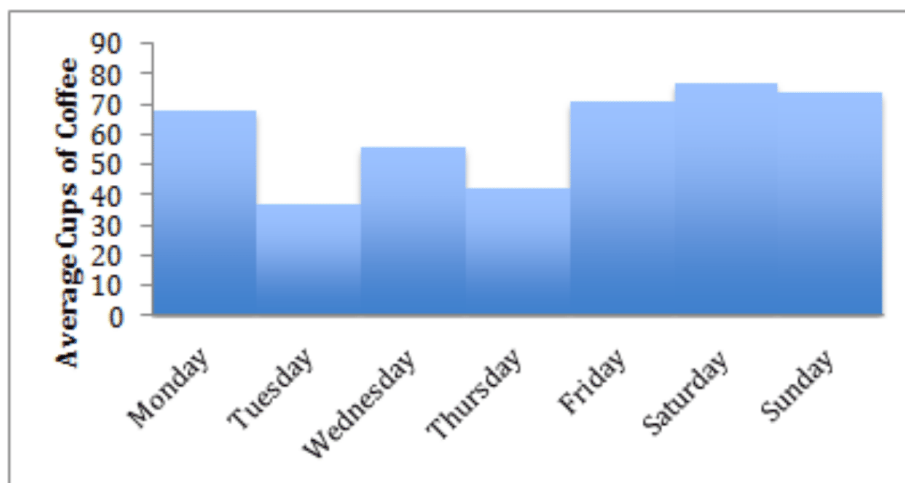
## 4.6 Interpreting Histograms

### Objective

Here you will learn how to read a histogram to develop conclusions based on data.

### Concept

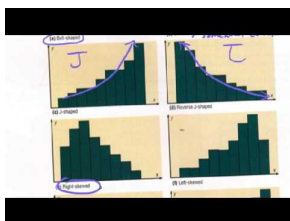
Brian runs a small business selling coffee to morning commuters. Like any other small business owner, he would like to increase his traffic. He decides to offer a “buy 10, get one free” punch-card promotion, and figures that if he gives double credit to customers for coffees purchased on days with generally slower sales, he can increase his overall traffic. If the histogram below represents the average number of coffees sold during a week, what can he conclude about the best and worst day(s) to offer “double punches”? How would you describe the shape of his histogram? Does the shape reveal any particularly useful data in this instance?



This lesson will help you become familiar with using histograms. At the end of the lesson, we will return to this question to apply your skills.

### Watch This

This is a good video describing the different shapes of data distributions. The instructor spends quite a bit of time at the beginning on the “bell curve” of a normal distribution, that we will be discussing in some detail later, so you may start at apx 3:30 if you wish to save time.



### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63707>

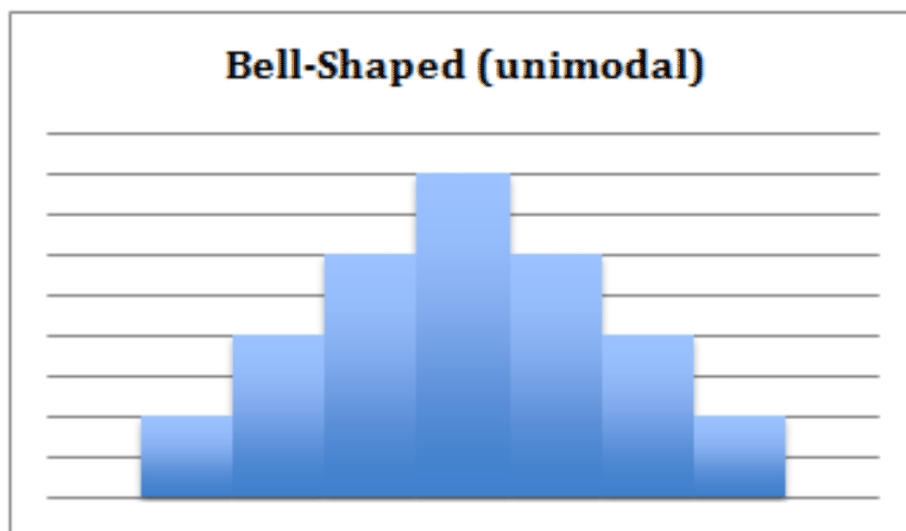
<http://youtu.be/LIT7qBmr bx A> mattemath - Types of Data Distributions (shapes and names)

## Guidance

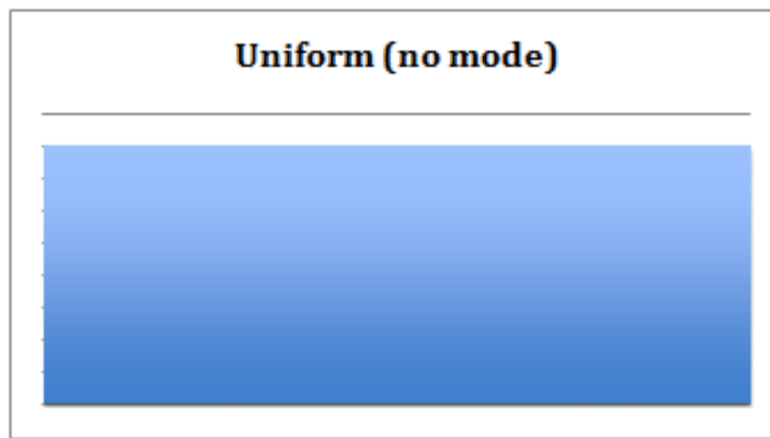
Histograms are a very common method of visualizing data, and that means that understanding how to interpret histograms is a valuable and important skill in virtually any career. There are a number of things to pay particular attention to when reading a histogram, including:

- **Range:** Recall that a range is a description of the difference between the greatest and least values in a given data set. On a histogram, this is important in two particular ways:
  - How widely dispersed are the frequencies of each bin? Extremely large frequency ranges (particularly as a percentage) may indicate data that is fundamentally unreliable.
  - How wide are the bins themselves? Specifically, how broad are the intervals or how descriptive are the classes? Unusually large or small intervals, or unusually broad or narrow categories may indicate important observations about the data as a whole.
- **Frequency Density:** The major difference between a bar graph and a histogram is the way in which the frequencies of each class or interval are represented. On a **bar graph**, the frequency is the *height* of the bar. On a **histogram**, the frequency is measured by the *area* of the bar. What that means is that you can use a histogram with different interval or class widths to represent data with varying densities. (See Example C)
- **Shape:** The shape of a histogram can lead to valuable conclusions about the trend(s) of the data. In fact, the shape of a histogram is something you should always note when evaluating the data the histogram represents. Some common shapes and their indications are:

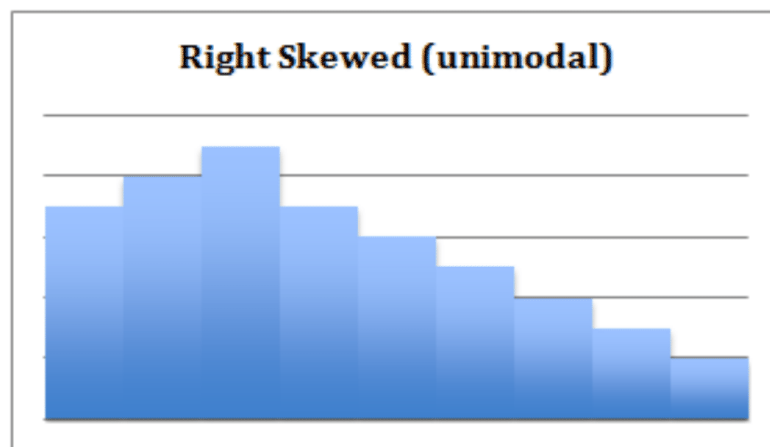
a. **Bell-Shaped:** A histogram with a prominent 'mound' in the center and similar tapering to the left and right. One indication of this shape is that the data is **unimodal** - meaning that the data has a single mode, identified by the 'peak' of the curve. If the shape is symmetrical, then the mean, median, and mode are all the same value. Note that a **normally distributed** data set creates a symmetric histogram that looks like a bell, leading to the common term for a normal distribution: a **bell curve**.



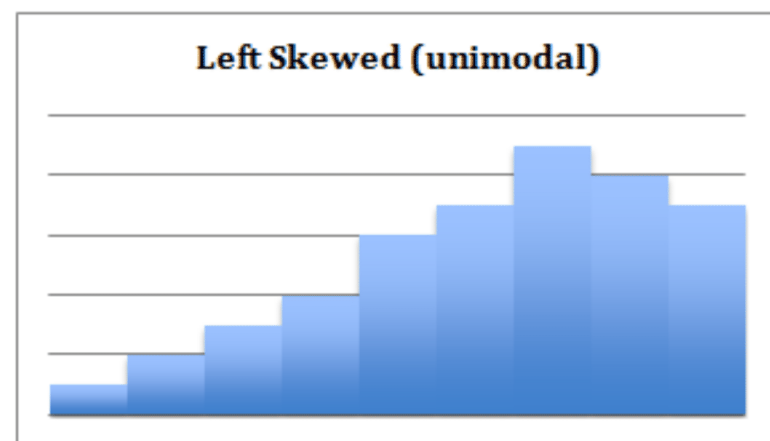
b. **Uniform:** A uniform shaped histogram indicates data that is very consistent; the frequency of each class is very similar to that of the others. A data set with a uniform-shaped histogram may be **multimodal** - the having multiple intervals with the maximum frequency. One indication of a uniform distribution is that the data may not be split into enough separate intervals or classes. Another possibility is that the scale of the histogram may need to be adjusted in order to offer meaningful observations.



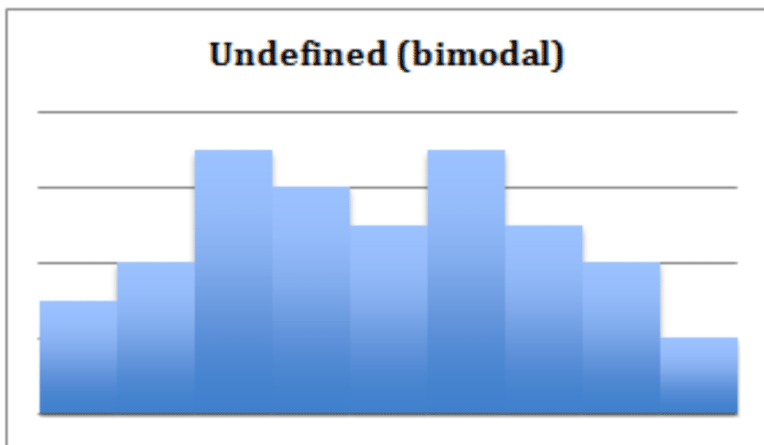
c. **Right-Skewed:** A right-skewed histogram has a peak that is left of center and a more gradual tapering to the right side of the graph. This is a unimodal data set, with the mode closer to the left of the graph and smaller than either the mean or the median. The mean of right-skewed data will be located to the right side of the graph and will be a greater value than either the median or the mode. This shape indicates that there are a number of data points, perhaps **outliers**, that are greater than the mode.



d. **Left-Skewed:** A left-skewed histogram has a peak to the right of center, more gradually tapering to the left side. It is unimodal, with the mode closer to the right and greater than either mean or median. The mean is closer to the left and is lesser than either median or mode. This shape indicates that the preponderance of any outliers is lesser than the mode.

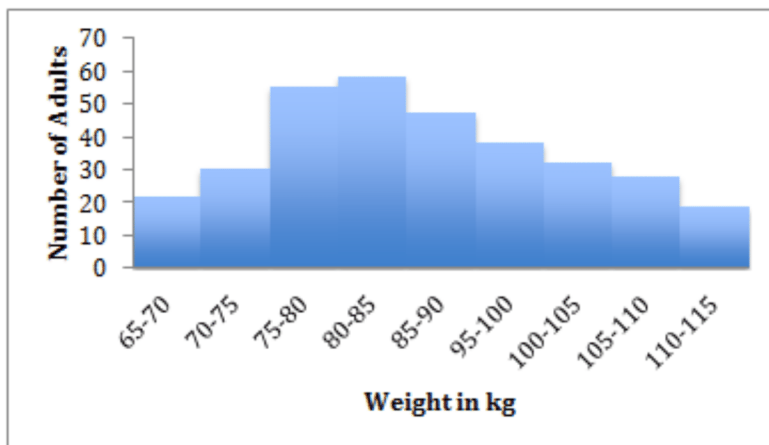


e. **Undefined Bimodal:** This shape is not specifically defined, but we can note regardless that it is bi-modal, having two separated classes or intervals equally representing the maximum frequency of the distribution.



### Example A

Describe the shape of the histogram and state a few notable characteristics:

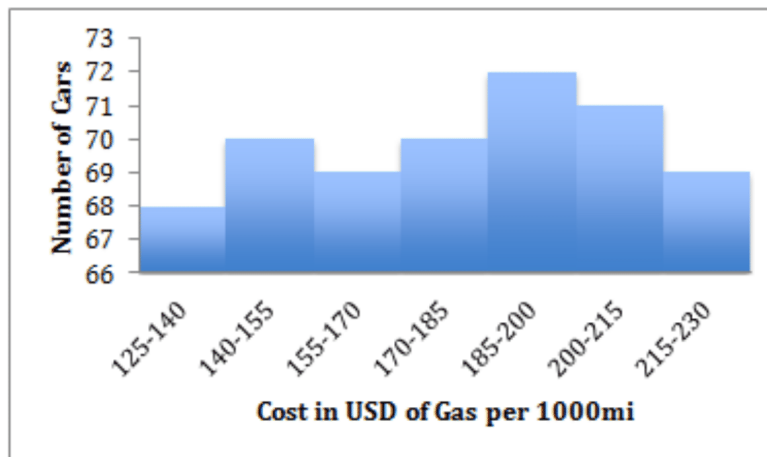


**Solution:** This is a right-skewed distribution, indicating that there are a number of values greater than the mode. If the modal class of 80-85kg represents a healthy normal weight, this graph would suggest a sample that tended toward being overweight.

### Example B

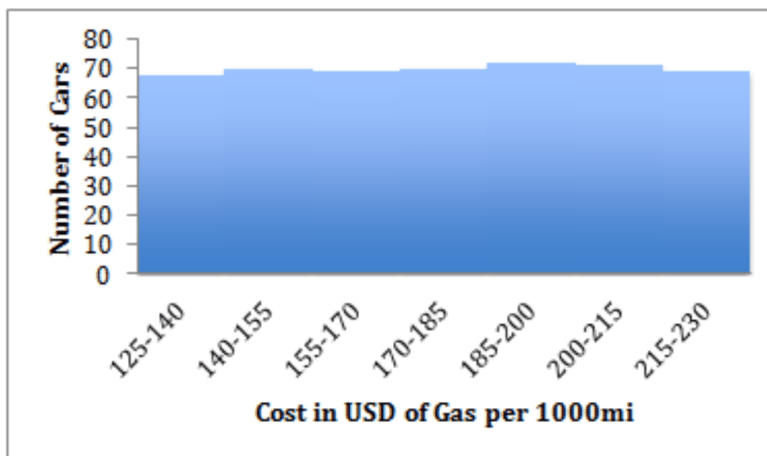
Identify the general shape of the histogram and what the shape indicates about the data:





**Solution:** This is a slightly tricky one. The overall shape appears somewhat left-skewed and obviously unimodal at first glance. However a closer look tells a different story, note that the overall range is  $72 - 68 = 4$  cars. That is a very small range, only about 5% of the mean. The shape is deceiving in large part because the vertical axis does not start at 0, which exaggerates the differences between the classes.

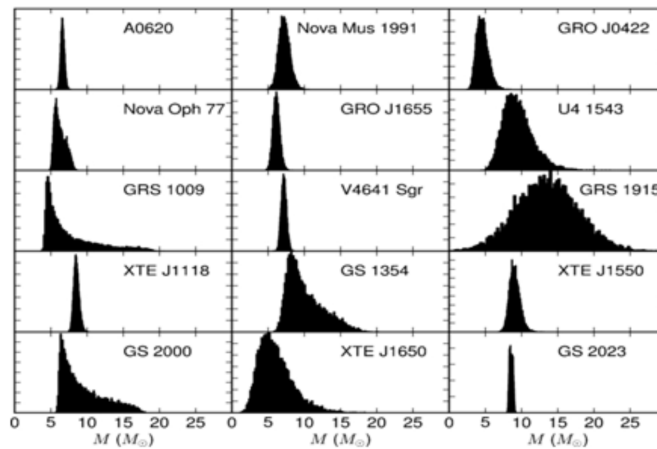
Look what happens if we re-draw the histogram *with the same data* but with the vertical axis at 0:



Pretty huge difference, isn't it? Now it is apparent that this is really a pretty uniform distribution, and that there is not a very meaningful difference in frequency between the classes.

### Example C

The image below represents data on the relative masses of a number of sampled black holes.



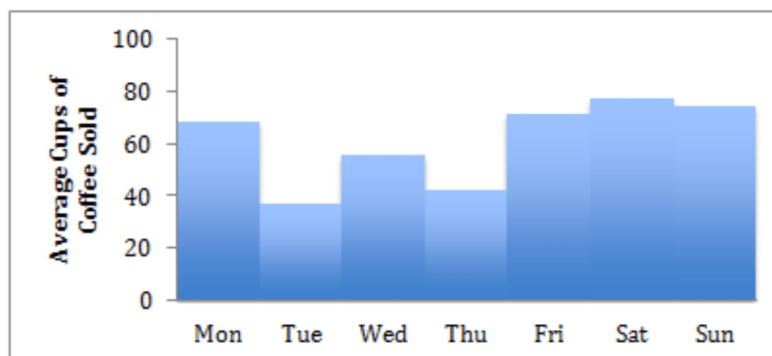
Evaluate the group of histograms as a whole; identify the common shapes and any notable features. Data source: <http://iopscience.iop.org>

**Solution:** Most of the individual histograms are clearly unimodal, and all are clustered rather closely around a single peak, with the exception of GRS 1915. Most of the graphs appear largely symmetrical, with the others being right-shifted. The sharp and narrow peaks in most of the plots suggest that the mass measurements are generally consistent. The location of the majority of the peaks at the same general location on the scale would suggest that the masses of the different black holes appear similar at this scale. The tendency of the non-symmetrical plots to be right-shifted suggests that it would be more reasonable to favor slightly greater mass estimates than slightly lesser ones.

The GRS 1915 plot is notably different, and the broad peak suggests that perhaps clear data on the mass of that particular black hole is difficult to come by.

### Concept Problem Revisited

*If the histogram below represents the average number of coffees sold during a week, what can he conclude about the best and worst day(s) to offer “double punches?” How would you describe the shape of his histogram? Does the shape reveal any particularly useful data in this instance?*



Brian should note that he is currently receiving less traffic on Tuesdays and Thursdays than he is the rest of the week. Those two days would be ideal for his “double punches”. This particular histogram does not have a well-defined shape, and therefore no particular information is liable to be pulled from it.

We might note that Friday, Saturday, Sunday, and Monday all in fact occur subsequently, so it could be said that the *data* suggests a peak during those days. As drawn, however, the histogram does not.

## Vocabulary

**Multimodal** histograms have more than one 'peak' in the data. Recall that the mode is the most common value, so a multimodal histogram represents data with multiple classes that have a frequency equal to the greatest single frequency in the data.

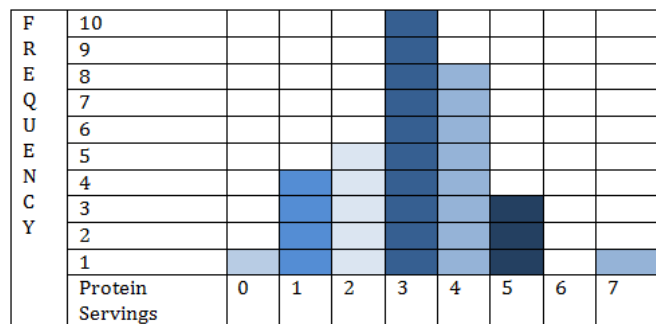
**Unimodal** histograms have a single peak, and represent data with a single most common frequency.

**Outliers** are uncommon frequencies occurring some distance from the peak. Specifically, data points that are more than 1.5 times the middle half of your data above the upper, or below the lower, quartiles may be considered **mild outliers** and points more than 3 times the middle half of your data are **extreme outliers**.

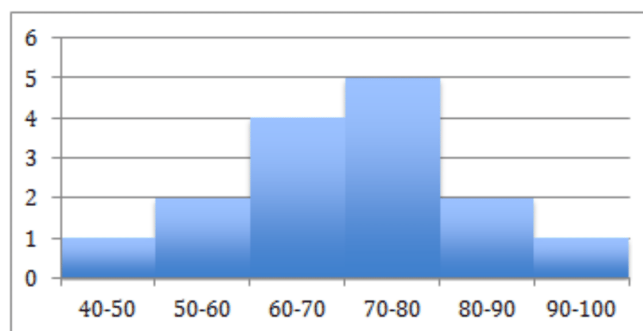
A **normal distribution** creates a histogram in the shape of a bell. This **bell curve** makes it clear that the majority of the data lies close to the mean.

## Guided Practice

A random sampling was taken of pedestrians on a street corner walking to work one day. They were asked how many servings of protein they ate, on average, on a daily basis. The data collected is displayed in the histogram below:



1. How many people say that they eat at least 4 servings of protein daily?
2. What percentage of people surveyed eat no more than 3 servings of protein daily?
3. What proportion of the people surveyed eats exactly 5 servings of protein daily?
4. What type of distribution does the histogram below display?



- a. Symmetric, single peaked (unimodal) distribution
- b. Symmetric, double peaked (bimodal) distribution
- c. Skewed left distribution

d. Skewed right distribution

5. Using the image from question 4, determine the spread and any outliers of this graph.

Solutions:

1. 8 people claim four servings per day, 3 claim five servings, and 1 claims seven servings, for a total of 12

2. 32 people responded, and of them 20 people eat 3 servings or less.

To find the percentage, divide the number who eat 3 or fewer servings by the total number of responses:  $\frac{20}{32} = 62\%$

3. 3 people claim five servings per day.

To find the percentage, divide:  $\frac{3}{32} = .094$  or  $\approx 10\%$

4. This is a symmetric, single peaked (unimodal) distribution.

5. No outliers.

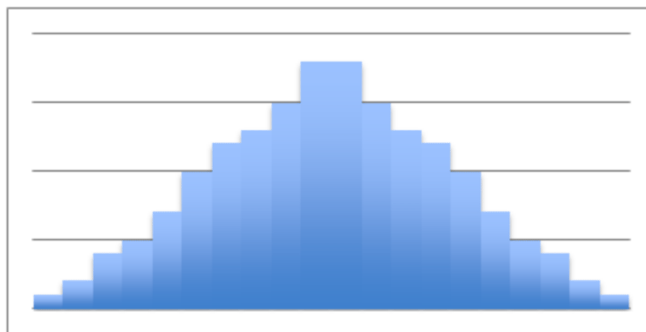
**TABLE 4.46:**

approximate min:	45 (the middle of the lowest interval of scores)
approximate max:	95 (the middle of the highest interval of scores)
approximate range:	$95 - 45 = 50$

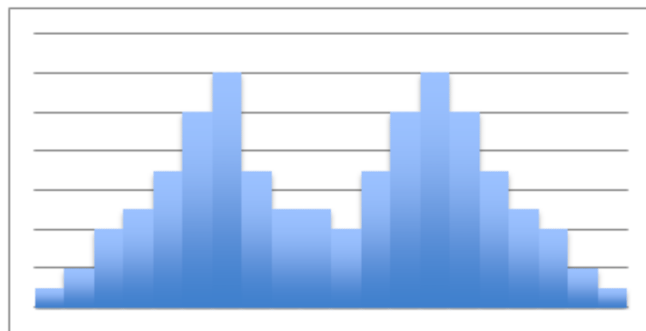
**Practice**

Identify which images show symmetric distributions and which show skewed distributions. Identify what type of symmetric or skewed distributions are displayed.

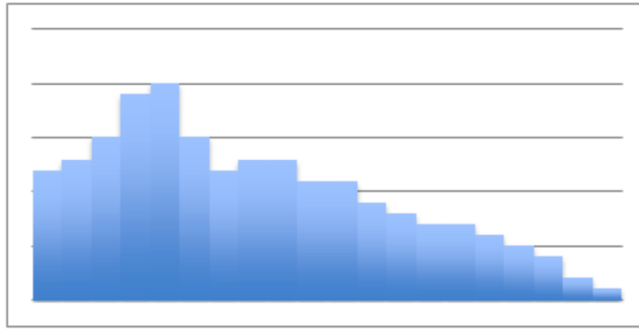
1.



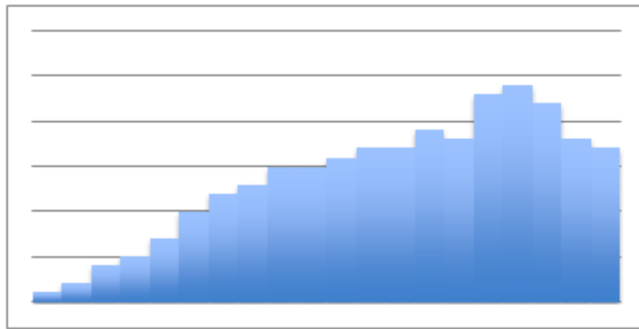
2.



3.



4.



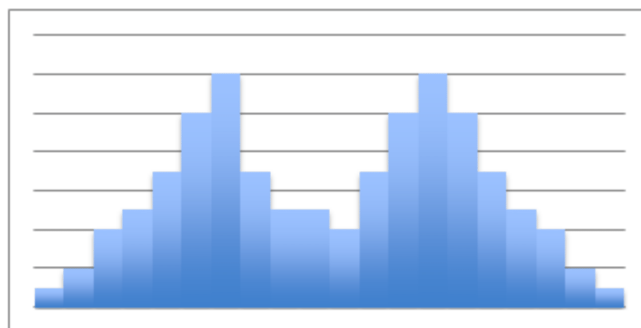
5. What do you think is the shape of the distribution of the age at which a child takes its first steps? Why?

- Symmetric - Uniform
- Skewed left
- Skewed right
- Symmetric - Unimodal
- Symmetric - Bimodal

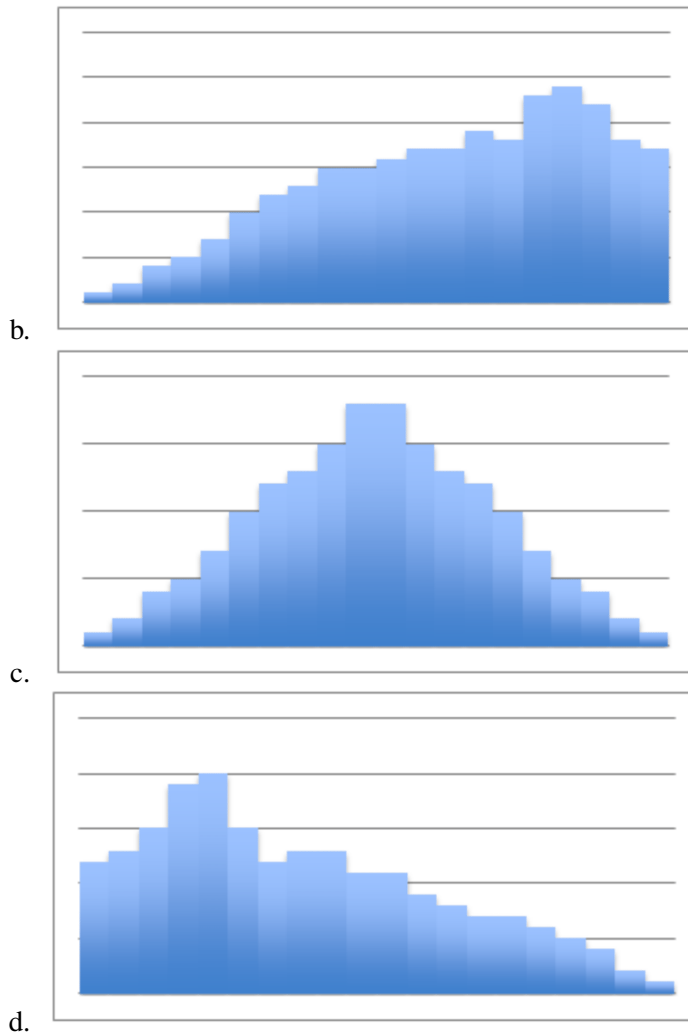
6. What do you think is the shape of the distribution of rolling a 6-sided die 1,000 times is? Why?

- Symmetric - Uniform
- Skewed left
- Skewed right
- Symmetric - Unimodal
- Symmetric Bimodal?

7. Match the graph with the data it most likely displays.



a.



SAT Math Scores of future doctors and engineers.

Prices of 1,000 homes within a given geographical area.

Cholesterol levels of 1000 adults.

Men's & women's clothing sizes.

The data below shows the number of surveyed people, and their respective ages, who enjoy riding roller coasters.

Use the histogram below to answer questions 8-11.

F R E Q U E N C Y	10											
	9											
	8											
	7											
	6											
	5											
	4											
	3											
	2											
	1											
	AGE	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-50	51-55	56-60

8. What is the shape of this histogram?

9. What is the center of this histogram?

10. What is the Spread of this histogram?

11. What are the outliers of this histogram?

Use the histogram below to answer questions 12-15.

70				█	
60				█	
50				█	
40			█	█	
30			█	█	
20		█	█	█	█
10	█	█	█	█	█
0	20	40	60	80	100

12. What is the shape of this histogram?

13. What is the center of this histogram?

14. What is the Spread of this histogram?

15. What are the outliers of this histogram?

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.6.

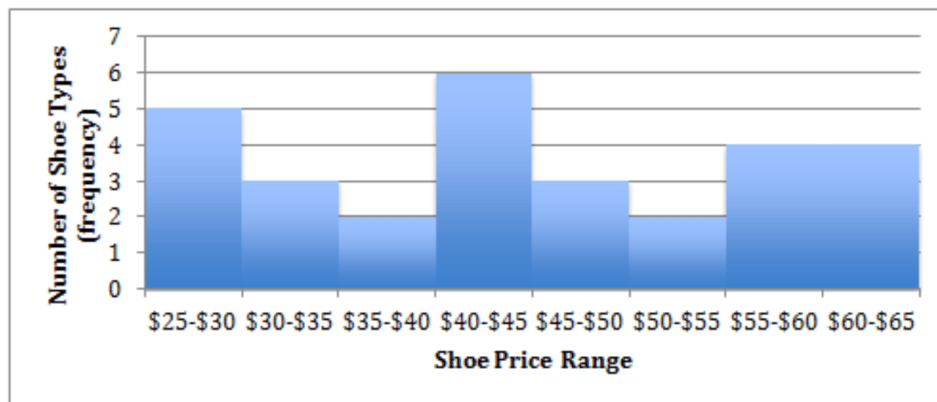
## 4.7 Frequency Polygons - Probability and Statistics

### Objective

Here you will learn to create frequency polygons and you will learn about the differences between frequency polygons and histograms or bar charts.

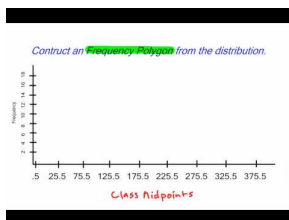
### Concept

The histogram below is built from data given in the lesson on histograms. How would you convert the histogram into an absolute frequency polygon or a relative frequency polygon?



After the lesson below, we will return to this question and apply what we have learned.

### Watch This



### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63714>

<http://youtu.be/gy4dgzxmT4A> Steve Mays - Frequency Polygon

### Guidance

A frequency polygon is closely related to a histogram. In fact, the difference is really only in the actual construction of the graph. Whereas a histogram is built of bins with a width representing the interval, and a height representing the quantity of data points in each interval, a frequency polygon is constructed by drawing a point to represent the



frequency of a particular interval and connecting that point to the one representing the frequency of the next interval. The result is a shape very much like a histogram constructed from the same data, but with points instead of columns.

The primary purpose of a frequency polygon is to allow histogram-like data representation of two sets of data on the same graph. Two histograms on the same graph tend to shroud each other and make comparison more difficult, but two frequency polygons can be graphed together with much less interference.

There are two common varieties of frequency polygon: *absolute frequency* and *relative frequency*. The difference between them is how the frequency is counted. An *absolute frequency* polygon has 'peaks' that represent the actual number of points in the associated interval. A *relative frequency* polygon has peaks that represent the percentage of total data points falling within the interval.

#### To create an absolute frequency polygon:

- Construct a frame just as you would for a histogram.
- Label the vertical axis with the range of frequencies to be graphed, and the horizontal axis with the intervals you have chosen. Make your horizontal axis long enough to include a full interval above and below your graphed data so that the finished polygon has a visible starting and ending point.
- Sum the number of points in each interval and mark a point representing the sum along the midline of the interval. The midline is on the arithmetic mean of the each interval, and can be calculated by adding the lower and upper limits of each interval and dividing the sum by two.
- Once all points have been accounted for, connect the points and color in the area under the line.
- If you are graphing a second set of data, repeat the process.

#### To construct a relative frequency polygon:

- Construct a frame just as you would for a histogram.
- Label the vertical axis from 0 - 100%, and the horizontal axis with the intervals you have chosen.
- Sum the number of points in each interval, divide the sum of each interval by the total number of data points, and multiply by 100. The result is the percentage of the total number of data points that is represented by each interval. Mark a point representing the percentage along the midline of the interval.
- Once all points have been accounted for, connect the points and color in the area under the line.
- If you are graphing a second set of data, repeat the process.

#### Example A

Construct an absolute frequency polygon to represent the data in the table.

**TABLE 4.47:** Time Spent Playing with Toys vs Unwrapping Presents and Eating on Christmas Day

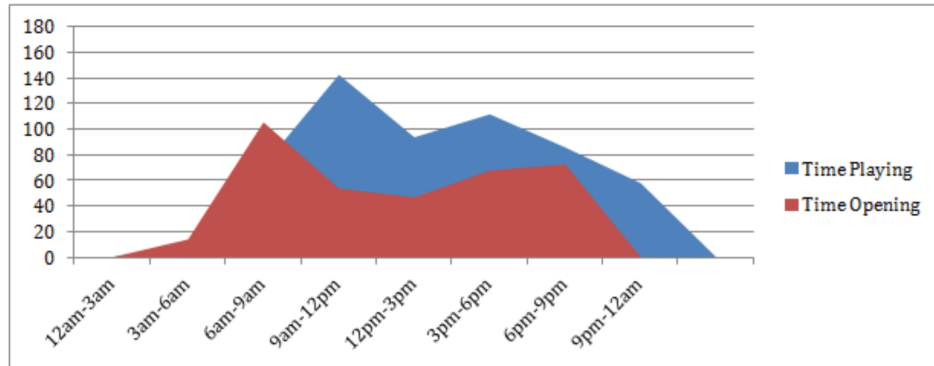
	12am - 3am	3am - 6am	6am - 9am	9am - 12pm	12pm - 3pm	3pm - 6pm	6pm - 9pm	9pm - 12am
Time Playing	0 min	10 min	72 min	143 min	94 min	112 min	86 min	58 min
Time Opening	0 min	14 min	106 min	54 min	47 min	68 min	73 min	0 min

**Solution:** First construct the frame for your graph; label the  $y$ -axis from 0 - 180 minutes, and the  $x$ -axis with each of the intervals described in the table (12am-3am, etc.).

Next, mark a single point along the midline of each bin, at the appropriate height to represent the frequency of minutes spent playing during that interval.

Finally, connect the points and color in the area between the  $x$ -axis and the line. Your completed frequency polygon

should look something like this:



### Example B

Construct a relative frequency polygon using the psychology test score data in the table.

**TABLE 4.48:**

Interval		Frequency
29.5	39.5	0
39.5	49.5	3
49.5	59.5	10
59.5	69.5	53
69.5	79.5	107
79.5	89.5	147
89.5	99.5	130
99.5	109.5	78
109.5	119.5	59
119.5	129.5	36
129.5	139.5	11
139.5	149.5	6
149.5	159.5	1
159.5	169.5	1
169.5	179.5	0

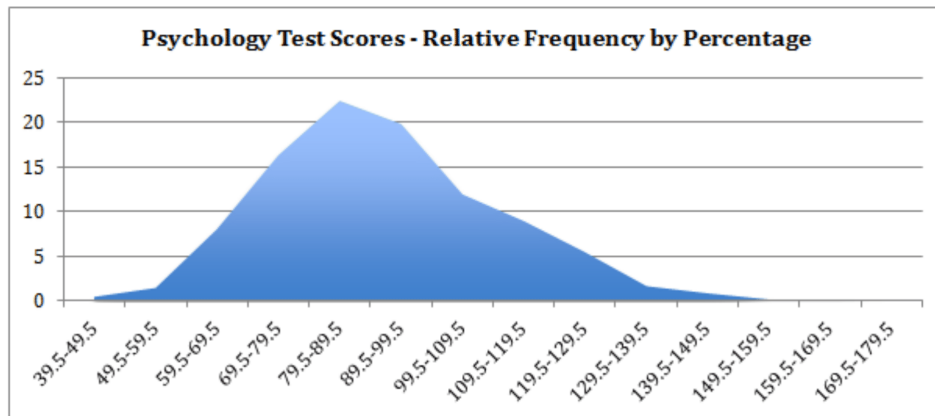
**Solution:** The data has been organized into intervals and the frequency for each has been noted. In order to convert the stated *absolute* frequencies into *relative* frequencies, first find the sum of the frequencies from all intervals and then divide each interval's frequency by the total frequency and multiply by 100 to get the percentage.

$$0 + 3 + 10 + 53 + 107 + 147 + 130 + 78 + 59 + 36 + 11 + 6 + 1 + 1 + 0 = 652 \text{ total scores}$$

- Relative frequency for the 1<sup>st</sup> interval, including scores between 29.5 and 39.5:
  - $\frac{0}{652} \times 100 = 0\%$
- Relative frequency for the 2<sup>nd</sup> interval, including scores between 39.5 and 49.5:
  - $\frac{3}{652} \times 100 \approx .5\%$  (round to 0%)
- Relative frequency for the 3<sup>rd</sup> interval, including scores between 49.5 and 59.5:
  - $\frac{10}{652} \times 100 \approx 1.5\%$

- Relative frequency for the 4<sup>th</sup> interval, including scores between 59.5 and 69.5:
  - $\frac{53}{652} \times 100 \approx 8\%$
- Repeat for the remaining values

Now you can create your graph as before, labeling the vertical axis from 1-25% (since none of the intervals are greater than 25% of the total) and the horizontal axis with the indicated intervals. The finished product should look rather like:



### Example C

The table below contains data on the ages of male and female students at a school dance. Create a multiple relative frequency polygon to illustrate the data in the table.

**TABLE 4.49:**

Interval (Age in Years)	Male Frequency	Female Frequency
13-13.5	6	3
13.5-14	5	6
14-14.5	9	13
14.5-15	13	19
15-15.5	12	16
15.5-16	16	17
16-16.5	14	11

**Solution:** First we will need to find the sum of each frequency column and divide the frequency of each interval by the sum to get the relative percentage of each interval. If we add a row for the sum of each column, and two columns for relative percentages, our revised table will look like this:

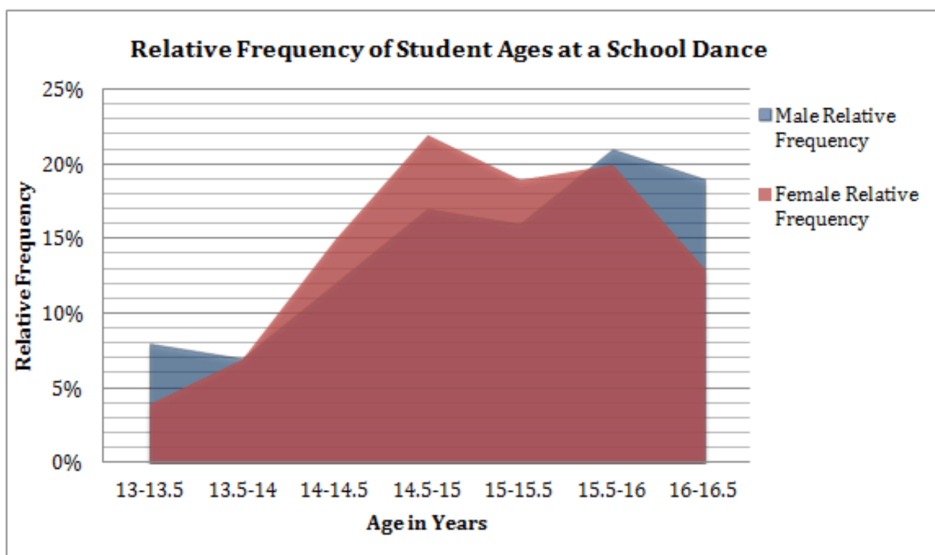
**TABLE 4.50:**

Interval (Age in Years)	Male Frequency	Male Relative Frequency	Female Frequency	Female Relative Frequency
13-13.5	6	8%	3	4%
13.5-14	5	7%	6	7%
14-14.5	9	12%	13	15%
14.5-15	13	17%	19	22%
15-15.5	12	16%	16	19%
15.5-16	16	21%	17	20%

**TABLE 4.50:** (continued)

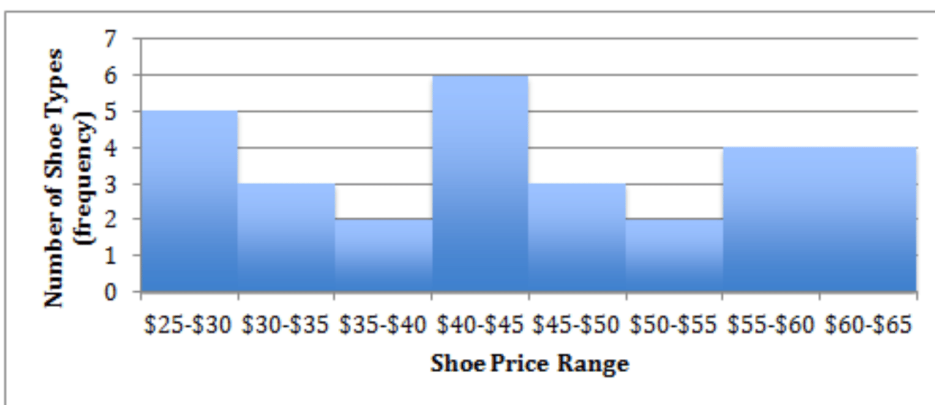
16-16.5	14	19%	11	13%
TOTAL	75	100%	85	100%

Then we need to graph both relative frequency polygons on the same graph so they can be easily compared.



**Concept Problem Revisited**

The histogram below is built from data given in the lesson on histograms. How would you convert the histogram into an absolute frequency polygon or a relative frequency polygon?



To convert the data into an absolute frequency polygon, plot a single point to represent the frequency of each interval at the midline of the interval, then connect the points with a line and color in the area between the line and the *x*-axis.

To convert into a relative frequency polygon, find the sum of frequencies of the entire sample, and divide the frequency of each interval by the sum, then multiply by 100 to get the relative percentage of each interval.

**Vocabulary**

*Absolute frequency* is an actual count of the number of data points that lie within a given interval.

**Relative frequency** is the ratio of the frequency of each interval to the frequency of the entire sample. May be expressed as a decimal between 0 and 1, or as a percentage.

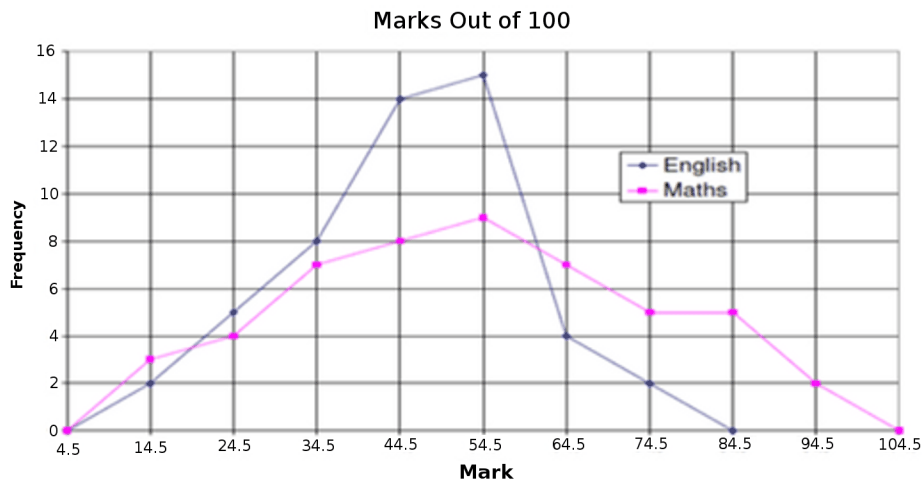
### Guided Practice

The following table represents the average times it took for 25 men and 25 women to run a race.

**TABLE 4.51:**

Time in Minutes	30-40	40-50	50-60	60-70	70-80
Frequency Men	0	4	3	8	10
Frequency Women	2	6	8	6	3

1. Create a frequency polygon (remember to use the midpoints of the time on your graph) displaying the above data.
2. Is the data continuous or discrete?
3. Look at the graph below. If the “Mark” represents the midpoint of the represented data, list the classes that the marks were grouped into.



Use the following data for Q's 4 and 5:

126, 146, 161, 156, 134, 164, 157, 156, 154, 156, 154, 156, 138, 164, 154, 173, 159, 143, 152, 132, 164, 158, 170, 149, 178, 136, 177, 165, 128

4. Complete the Table

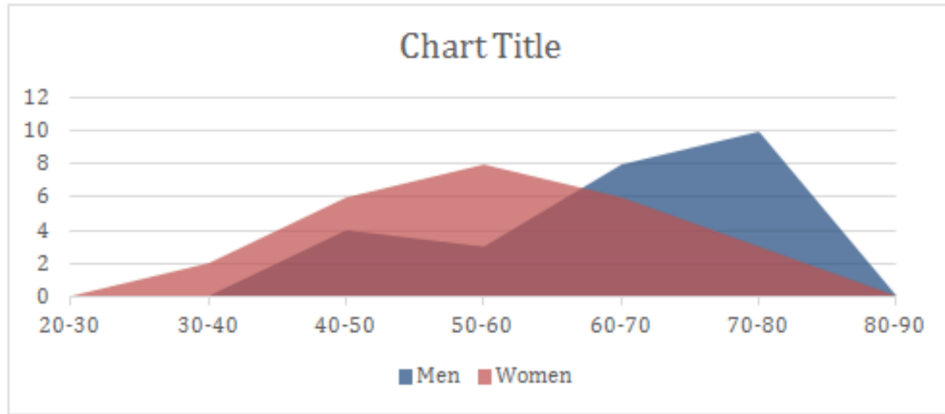
**TABLE 4.52:**

Class Interval	Midpoint	Frequency
120-129		

5. Create an Absolute Frequency Polygon for the data

Solutions:

1. Build a chart with intervals along the bottom (on the x-axis) matching those in the table with one category above and below, and with frequency values from 1- 10 up the side (y-axis):



2. The data is continuous, divided into intervals.

3. Remember that the data point is the *center* of the interval it represents, so the intervals are:

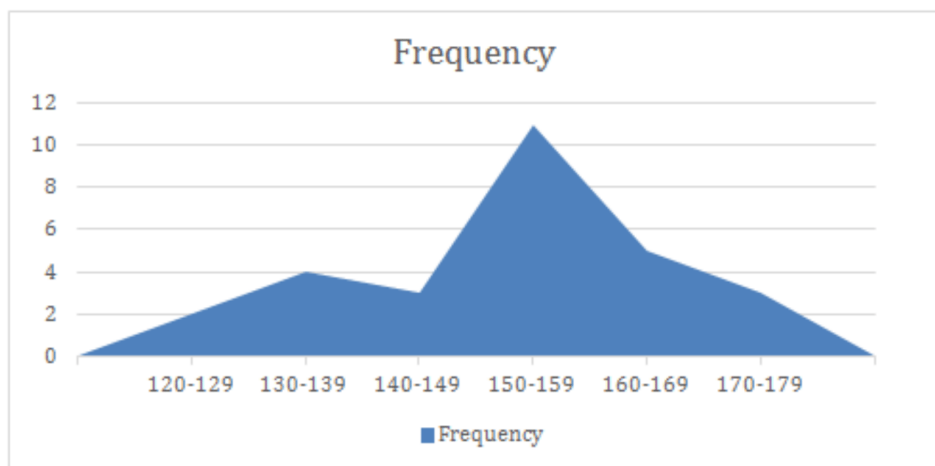
0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99, 100-109

4. The table shows an initial interval of 10 units, filling in the remainder of the intervals at the same range yields:

**TABLE 4.53:**

Class Interval	Midpoint	Frequency
120-129	124.5	II
130-139	134.5	III
140-149	144.5	III
150-159	154.5	IIIIIIIIII
160-169	164.5	IIII
170-179	174.5	III

5.



**Practice Questions:**

1. Why is a frequency polygon called a “polygon”?

2. What is the advantage of a frequency polygon?

A police radar unit measured the speed of 25 cars on a certain street during the week.

The resulting speeds were:

29, 23, 30, 30, 27, 24, 30, 25, 23, 28, 25, 24, 28, 30, 23, 30, 27, 25, 29, 24, 23, 26, 30, 28, 25

The policeman then completed the same procedure at the same time of day, only on a weekend.

The resulting speeds were:

30, 35, 25, 27, 30, 26, 25, 32, 29, 30, 28, 27, 27, 34, 32, 30, 27, 26, 31, 32, 35, 29, 33, 32, 28

3. Create a frequency table for both sets of data, assume an interval of 2 mph.

4. Create a combined absolute frequency polygon for the above data (both sets on the same chart).

5. What information can be derived from the two sets of data? What can you infer about weekday vs. weekend drivers?

6. If you had to take an educated guess, what would you say the speed limit on the street is?

Based on the following data:

11, 21, 30, 40, 57, 37, 51, 27, 35, 47, 33, 50, 14, 23, 63, 42, 66, 38, 65, 19, 43, 31, 32, 53, 56, 33, 25, 44, 52, 39

7. Complete the Table

**TABLE 4.54:**

Class Interval	Midpoint	Frequency
10-19		

8. Create an Absolute Frequency Polygon for the data.

Based on the following data:

45, 70, 58, 71, 63, 65, 88, 69, 71, 62, 75, 89, 75, 67, 70, 83, 73, 98, 77, 91, 52, 65, 87, 95, 43, 97, 53, 71, 66, 88

9. Complete the Table

**TABLE 4.55:**

Class Interval	Midpoint	Frequency
40-49		

10. Create an Absolute Frequency Polygon for the data.

Based on the following data:

108, 129, 133, 127, 117, 148, 128, 144, 130, 144, 120, 123, 103, 134, 152, 124, 134, 122, 147, 125, 117, 131, 148, 128, 134, 106, 159, 154, 121, 138

11. Complete the Table

**TABLE 4.56:**

<b>Class Interval</b>	<b>Midpoint</b>	<b>Frequency</b>
100-109		

12. Create an Absolute Frequency Polygon for the data.

13. Create a Relative Frequency Polygon for the data, and compare the appearance with that of the absolute frequency polygon.

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.7.



## 4.8 Creating Box-and-Whisker Plots

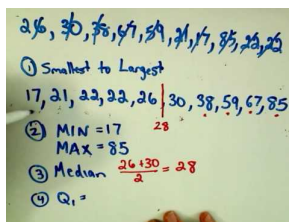
### Objective

Here you will learn how box-and-whisker plots are created, and some common uses of them.

### Concept

If you were asked to create a visual representation of the mean, upper and lower 25% (quartiles), and maximum and minimum (extremes) scores on the final test in your College Algebra class, how would you go about it? Would a box-and-whisker plot be appropriate? Why or why not? What would the plot look like if the mean was 82%, the lowest score was 59%, highest was 96%, and if a quarter of the class scored above 86% while another quarter scored below 70%?

### Watch This



### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63751>

<http://youtu.be/GMb6HaLXmjY> PatrickJMT - Box and Whisker Plot

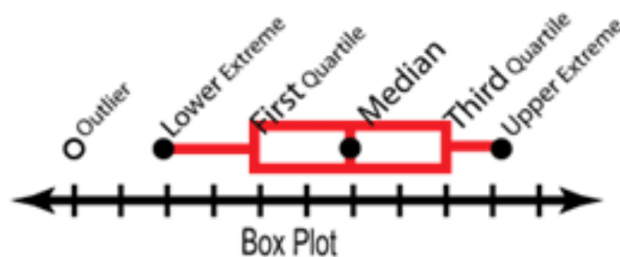
### Guidance

**Box-and-whisker plots** (or **box plots**) are ideal for visually representing the *five number summary* of data.

First, *organize the data by increasing value*, then

**The five number summary (or five statistical) is composed of:**

- The minimum and maximum values - called the *extremes*
- The middle value - called the *median*
- The values halfway between each extreme and the median - called the *quartiles*.



It is important to recognize that a five number summary is more dependent on the *position* of each value in numerical order than on the value itself. A common confusion when gathering data for a box plot is to think that the plot is based on the *arithmetic mean* of the data rather than the median, don't fall into this trap! To create a box plot, which is based on the five number summary, you first need to organize your data in increasing numerical order, and then identify your five numbers *based on position in the ascending series*.

A sometimes tricky detail is the handling and identification of *outliers*. Once you have identified the median and quartiles of your data, you should review the values at the lower and upper limits to see if there are any that seem unusually extreme before considering them to be part of your 5 number summary. Specifically, data points that are more than 1.5 times the *inter-quartile range* (the range of values between the first and third quartiles - representing the middle half of your data), may be considered *mild outliers*. Any points more than 3 times the inter-quartile range may be considered *extreme outliers*. Outliers are commonly plotted as stars or asterisks (mild outliers) or open circles (extreme outliers), and are not a part of the actual box plot or the five number summary.

Once you have identified your five number summary, create a number line extending at least 10% past the upper and lower extremes of your data, and plot each of the five numbers above the appropriate locations on the line. Now create a rectangular box with sides on the 1<sup>st</sup> and 3<sup>rd</sup> quartiles. Draw a vertical line inside the box to represent the median, and draw horizontal lines from the sides of the box to the extremes. Finally, identify any mild outliers with asterisks/stars and extreme outliers with open circles.

### Example A

Create a five number summary for the data below, and identify any outliers:

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

**Solution:** A five number summary includes the median, the upper and lower extremes, and the first and third quartiles. The first step to identifying them is to organize the data by ascending numerical value:

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

- Finding the median: Note that there are 15 values, an odd number, so the middle number in the series is the median. The value "5" has 7 values above and 7 below. **5 is the median.**
- The 1<sup>st</sup> quartile is the median of the lower half of the data. There are 7 values below the median, and the middle number of them is "2", with three values below and three above before the median. **The 1<sup>st</sup> quartile is 2.**
- The 3<sup>rd</sup> quartile is the median of the upper half of the data. There are 7 values above the median, and the middle value is "7", with three values above it and three below it before the median. **7 is the 3<sup>rd</sup> quartile.**
- Are there any outliers? The inter-quartile range is the difference between the 1<sup>st</sup> and 3<sup>rd</sup> quartiles:  $7 - 2 = 5$ . Recall that a value should be considered an outlier if it is unusually low in frequency and greater than 1.5 times the inter-quartile range from the median. In this case, that would mean any number more than 7.5 above the 3<sup>rd</sup> quartile, 7, or below the 1<sup>st</sup> quartile, 2. That would make any value less than -5.5 or greater than 14.5 be considered a mild outlier. There are no negative values and no values greater than 9, so **there are no outliers.**
- The minimum and maximum values are the least and greatest values, respectively. Since we have organized our data in ascending order, the minimum value is on the far left, "1", and the maximum value is on the far right, "9". **The minimum is 1 and the maximum is 9.**

### Example B

Identify the five statistical summary and any outliers in the data below:

18, 16, 18, 17, 15, 2, 17, 20, 19, 18, 15, 16, 28, 18

**Solution:** Begin by ordering the data numerically:

2, 15, 15, 16, 16, 17, 17, 18, 18, 18, 18, 19, 20, 28

- Median: There are 14 values, an even number, so the median is the average (arithmetic mean) of the two

middle numbers, 17 and 18. **17.5 is the median.**

- **1<sup>st</sup> and 3<sup>rd</sup> quartiles:** The middle number in the lower 50% is 16, and the middle of the upper 50% is 18. **16 is the lower quartile and 18 is the upper quartile.**
- The inter-quartile range is  $18 - 16 = 2$ . Any value less than 13 or greater than 21 may be considered a mild outlier, and any value less than 10 or greater than 24 may be considered an extreme outlier. **2 and 28 are both extreme outliers.**
- The least value is 2 and the greatest value is 28. **2 is the minimum and 28 is the maximum.**

### Example C

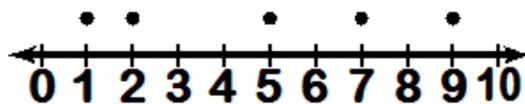
Create box plots representing the data from examples A and B.

Solution:

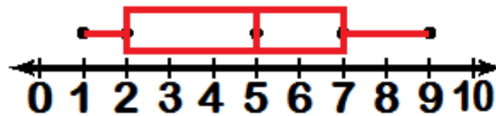
A. The data from example A was encapsulated in the five number summary:

Median: 5    1<sup>st</sup> quartile: 2    3<sup>rd</sup> quartile: 7    Minimum: 1    Maximum: 9

Draw a number line running from 0 to 10, and plot the five number summary above it:



Draw a rectangle including the first and third quartiles, and a vertical line for the median. Since there are no outliers, draw a “whisker” from each side of the box to the extremes:



B. The data from example B includes:

Median: 17.5

1<sup>st</sup> quartile: 16                      3<sup>rd</sup> quartile: 18

Minimum: 2                          Maximum: 28

Outliers (extreme): 2, 28

Draw a number line running from 0 - 30, and plot the five number summary:



Note that since “2” and “28” are both extreme outliers, the box-and-whiskers only extend to the greatest and least non-extreme values. This is sometimes called a **modified boxplot**.

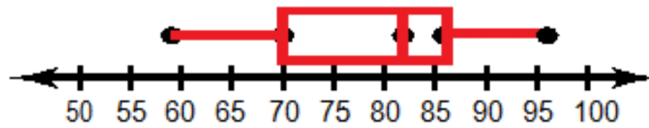
### Concept Problem Revisited

If you were asked to create a visual representation of the mean, upper and lower 25% (quartiles), and maximum and minimum (extremes) scores on the final test in your College Algebra class, how would you go about it? Would a box-and-whisker plot be appropriate? Why or why not? What would the plot look like if the mean was 82%, the

lowest score was 59%, highest was 96%, and if a quarter of the class scored above 86% while another quarter scored below 70%?

This would be an excellent application for a box plot. In fact, this is just about the *best* use of one. You will find, if you haven't already, that the SAT and ACT college application exams report grades in just this manner. Colleges (and students themselves) inevitably wish to see how a particular score compares to others on the same test, and a box plot is ideal for that purpose.

If the data in the question were plotted as a box plot, it would appear like this:



## Vocabulary

The **5 number summary** (or 5 statistic summary) is the collective term used to describe the minimum and maximum, middle, and 25% and 75% values in a data set.

The **extremes** are the minimum and maximum values in a set of data.

The **median** is the middle value in a set of data, when the data is organized in numerical order.

The values halfway between each extreme and the median are called the **quartiles**.

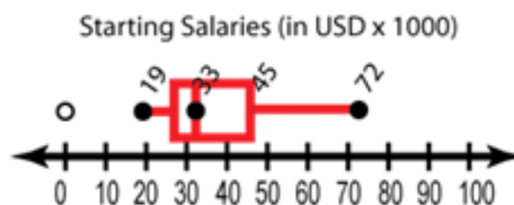
The **arithmetic mean** is a measure of central tendency calculated by finding the sum of the data, divided by the number of data entries. This value is referred to as the **average** in common language.

The **inter-quartile range** is the range of values between the first and third quartiles - representing the middle half of your data. In other words: the  $IQR = Q3 - Q1$ , and 50% of the data is in the box.

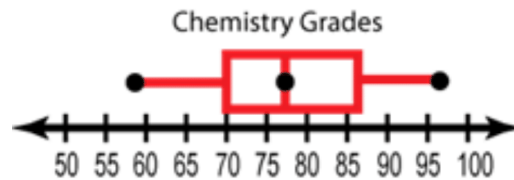
**Outliers** are values uncommonly distant from the mean. Mild outliers are determined as values at least 1.5 times the inter-quartile range above or below the 3<sup>rd</sup> or 1<sup>st</sup> quartiles. Extreme outliers are values greater than 3 times the inter-quartile range from the upper or lower quartiles.

## Guided Practice

1. The box-and-whisker plot below shows the starting salaries for graduates of a small college. What is the range of the starting salaries?



2. Mr. Andrews made a box-and-whisker graph of the quiz grades in his chemistry class. What is the median quiz grade for the class?



3. Mr. Foreman grades on a curve in which the top 25% of the test scores earn A's, the middle 50% earn C's, and the bottom 25% earn F's. The box and whisker plot below shows the distribution of scores on the last test. What is the range of scores for people who earned C's?



### Solutions:

1. The range is the difference between the maximum and minimum values, that is:

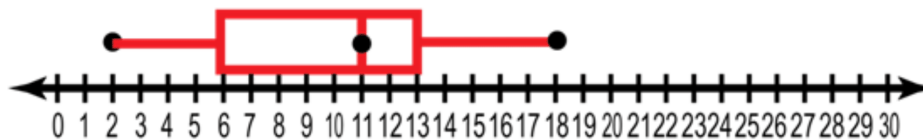
$$\$72,000 - \$19,000 = \$53,000$$

2. The median is denoted by a line in the center of a boxplot, in this case, that would be **77%**

3. If the middle 50% of Mr. Foreman's class earns C's, then all of the scores in the interquartile range, the area between Q1 and Q3, would be included. Since the "box" of a boxplot indicates the IQR, that would be **65% - 80%**.

### Practice

Use the following boxplot to answer questions 1 - 5 below:



1. What is the median
2. What is the lower quartile
3. What is the upper quartile
4. What is the minimum value
5. What is the maximum value
6. What are the five values called?
7. What is the range of the data?
8. What percentage of the data is below the upper quartile?
9. What percentage of data is located between the lower quartile and the median?
10. What percentage of data is above the median?
11. What percentage of data is below the lower quartile?
12. Calculate the Range for the following data: 5, 21, 10, 9, 12, 12, 16, 16, 9, 6, 20, 8, 10, 26, 4, 26, and 14.
13. Calculate the First Quartile for the following data: 5, 21, 10, 9, 12, 14, 13, 16, 9, 6, 20, 8, 12, 24, 4, 26, and 14
14. State the five number summary of the following data set: 13, 14, 10, 4, 18, 17, 11, 10, 5, 7, 10, 19, 13
15. Construct a box and whisker plot for the data set given in question 12

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.8.

## 4.9 Interpreting Box-and-Whisker Plots

### Objective

Here you will learn to efficiently pull information from box plots.

### Concept

If you were asked to evaluate a box plot to find the median, quartiles, extremes and outliers, would you know how? What does it mean if the 'box' in a box plot is unusually long or short? Does a long 'whisker' on one or both sides mean something important?

Review the lesson below and we'll return to answer these questions at the end.

### Watch This



#### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/62550>

<http://youtu.be/b2C9I8HuCe4> Khan Academy - Reading Box and Whisker Plots

### Guidance

Box-and-whisker plots (or “box plots”) are commonly used to compare a single value or range of values for easier, more effective decision-making. Box and whisker plots are very effective and easy to read, and can summarize data from multiple sources and display the results in a single graph.

Use box and whisker plots when you have multiple data sets from independent sources that are related to each other in some way. Examples include comparing test scores between schools or classrooms, and exploring data from before and after a process change.

Remember that the line inside the box represents the middle value when the data points are arranged numerically. Because the median is only identified by location in a series, it can sometimes be very indicative of the trend or average of the data set as a whole, and sometimes is not useful for that purpose at all (see Example A).

Recall that skewed data appears as a longer “tail” in one direction on a histogram, it is similar on a box plot. If the box in a box plot is stretched in one direction or the other, then the data is skewed in that direction. Data skewed right indicates a closer concentration of values on the *left*, since the plot indicates values more “strung out” on the right side.

A longer box indicates a greater *interquartile range* since the sides of the box indicate the 1<sup>st</sup> and 3<sup>rd</sup> quartiles. A greater interquartile range is an indicator of data that may be somewhat unreliable. Since the interquartile range

represents the 50% of the data closest to the median, a greater range in this section of the plot suggests that the median may not be a great indicator of central tendency.

A plot with long whiskers represents a greater range for the overall sample than simply a longer box itself does. Data covering a greater range is naturally less reliable as an indicator of highly probable values, but given the option, longer whiskers are less of a concern than a long box. A broad range of possibilities but a strong likelihood of central values is more reliable to use for prediction than a moderate overall range with little concentration at the median.

**Example A**

Identify the 5 number summary and any outliers depicted in the box plot below:

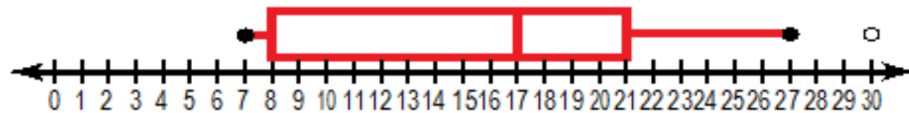


**Solution:** The 5 number summary is depicted by the vertical bars in the box and by the endpoints of the 'whiskers':

- Minimum: 13
- 1<sup>st</sup> Quartile: 16
- Median: 19
- 3<sup>rd</sup> Quartile: 22
- Maximum: 24
- Outliers (depicted by open circles disconnected from the box and whiskers): 4 and 30

**Example B**

What is indicated by the shape of the box plot below?

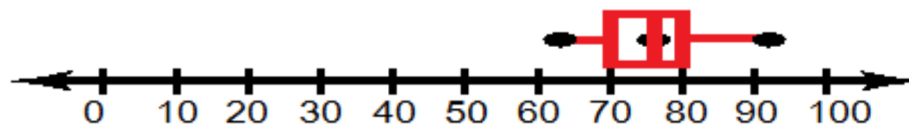


**Solution:** The box in the plot extends nearly to the lower extreme, indicating that the data less than the median is likely at least relatively consistent, since there is not a large jump between the lower 25% and the minimum. The longer whisker on the upper side suggests that there may be larger variance among the greater values, since there is a greater distance from the 3<sup>rd</sup> quartile to the upper extreme than from the median to the 3<sup>rd</sup> quartile.

**Example C**

A percentile box plot compares a particular value or range of values to an averaged reference point. The values on the scale represent the percentage of scores less than the plotted value. For instance, a score of 55% indicates that 55% of other values were less than the indicated score, and 45% were greater.

Maria recently completed a standardized test, and the box plot below describes her results. The median is her actual calculated percentile, and the rest of the 5 number summary suggests the range of percentiles that her score is expected to lie within once all scores are tabulated. Based on the information in the graph, would you expect Maria to be proud of her score? Why or why not?





**Solution:** Maria's score is expected to lie between the 62<sup>nd</sup> and 92<sup>nd</sup> percentile, with the most likely comparison being the 76<sup>th</sup> percentile. Since the 76<sup>th</sup> percentile indicates that her score was higher than that of 76% of all the students who took the test, and only 24% achieved a higher score than hers, yes, I would certainly say she has reason to be proud!

### Concept Problem Revisited

If you were asked to evaluate a box plot to find the median, quartiles, extremes and outliers, would you know how? What does it mean if the 'box' in a box plot is unusually long or short? Does a long 'whisker' on one or both sides mean something important?

With the practice you have had now, these questions should be easy!

- Median: the center vertical line in the 'box'
- 1<sup>st</sup> and 3<sup>rd</sup> Quartiles: the leftmost and rightmost vertical lines of the 'box'
- Lower and Upper Extremes: the endpoints of the 'whiskers'

### Vocabulary

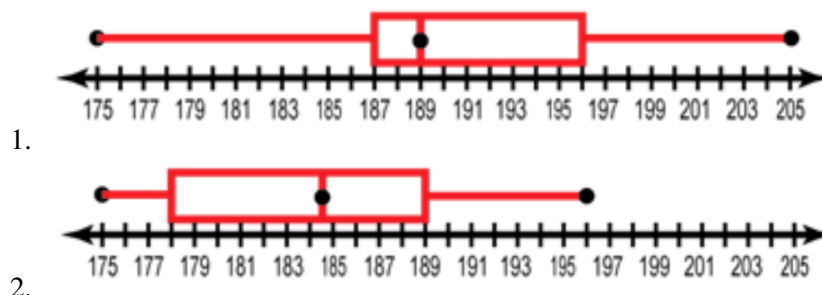
The *interquartile range* is calculated by subtracting the 1<sup>st</sup> quartile from the 3<sup>rd</sup> quartile and represents the middle 50% of the sample.

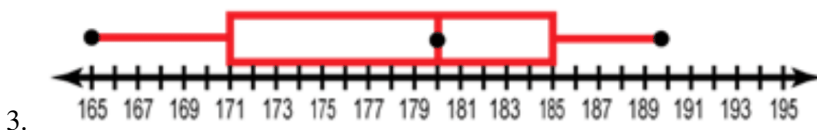
### Guided Practice

1. Make a Box and Whisker plot from the following data sets.
  1. Initial weight (December) of 14 women in a weight loss study (pounds) 190, 175, 187, 199, 205, 187, 176, 180, 187, 191, 200, 193, 188, 196
  2. Weights of the same women one month later (January) 187, 174, 181, 189, 196, 178, 174, 176, 181, 186, 188, 191, 183, 191
  3. Weights of the same women in February. 181, 165, 176, 182, 190, 176, 171, 170, 171, 185, 187, 181, 179, 186
2. How do the data in a and c compare?
3. How did the median change?
4. How did the maximum weight change?
5. How did the minimum weight change?
6. How did the range change?
7. How would you judge the effectiveness of the weight loss method used in the study?

### Solutions:

1. For all three sets, first organize the data by increasing numerical order and identify the five-number summary (FNS). Once you have the FNS, create the box plot for each just as in the examples above. The three plots should resemble the images below:

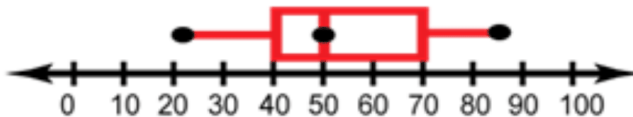




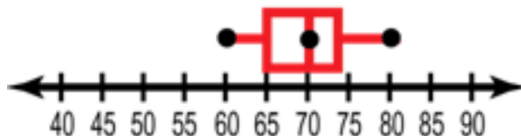
- 3.
2. If we compare the data between a and c, we can see that the overall weights of the women in the study did indeed go down. In fact, the minimum value at the start of the study was greater than the maximum two months later.
3. The median in December was 189, and in February it was 180.
4. The maximum in December was 205, and went down to 190 by February.
5. The minimum weight in December was 187, and it also went down, to 171 by February.
6. The range increased notably, from a mere 9 pounds in December, to more than 1.5 times that, 14, in January.
7. It would appear that the method was effective, at least in the short term. The increased range would indicate that it was somewhat more effective for some participants than others.

**Practice**

1. What is the five number summary of the following box and whisker plot?



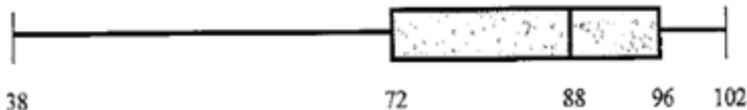
2. The box plot shows the heights in inches of boys on a High School Baseball Team. What is the 5 number summary of the plot?



3. Listed are the heights in inches of girls on a High School Ski Team. Make a plot of the girls' heights. 58, 59, 59, 60, 62, 65, 68, 69, 70, 70, 71

4. Comparing the heights between the two teams, which has the taller players on average? How do you know?

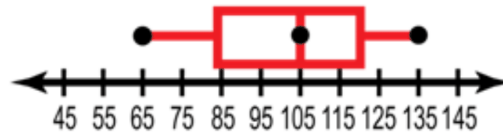
Use the box and whisker plot below to examine scores received on an English GED Test to answer questions 5-9



5. What was the high score on the test?
6. What percent of the class scored above a 72?
7. What was the median score on the test?
8. What percent of the class scored between 88 and 96?

9. Would you expect the mean to be above or below the median? Explain

Use the graph below that shows how much girls spent on average per month on clothes during August.



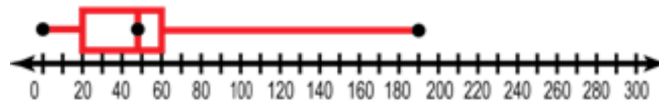
10. How many girls shop for clothes?

11. What percent of girls spent less than \$85.00 in August on clothes?

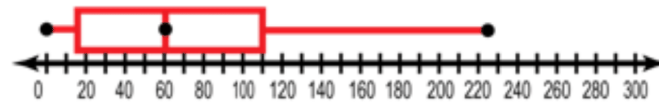
12. Would you expect the mean number of dollars spent to be higher or lower than the median? Explain

Use the graphs below to compare the amount of time a teenager spends in the bathroom getting ready for school and the amount of time they spend in the bathroom getting ready to go to a party.

TIME SPENT GETTING READY FOR SCHOOL:



TIME SPENT GETTING READY FOR A PARTY:



13. What percent of teenagers spend at least 15 minutes getting ready for a party?

14. What is the 3<sup>rd</sup> Quartile for the time spent getting ready for a party?

15. Is it more common for a teenager to spend more than 1 hour getting ready for school or between 1 and 2 hrs getting ready for a party? Explain

Answer True or False for questions 16-24.

16. \_\_\_\_\_ Some teenagers do not spend time getting ready for parties.
17. \_\_\_\_\_ The graph of time spent getting ready for a party contains more data than the getting ready for school graph.
18. \_\_\_\_\_ 25% of teenagers spend between 48 and 60 minutes getting ready for school.
19. \_\_\_\_\_ 15% of the teenagers did not go to parties that month
20. \_\_\_\_\_ In general teenagers spend more time getting ready for a party than getting ready for school.
21. \_\_\_\_\_ The Party data is more varied than the homework data
22. \_\_\_\_\_ The ratio of teenagers who spend more than 110 minutes getting ready for a party to those who spend less is about 2:1
23. \_\_\_\_\_ 225 Teenagers watch TV.
24. \_\_\_\_\_ Twice as many teenagers spend more than 1 hour on getting ready for school, than they do spending an hour getting ready for a party.

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.9.

## 4.10 Creating Stem-and-Leaf Diagrams

### Objective

Here you will learn how to create Stem-and-Leaf Diagrams.

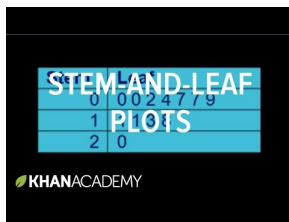
### Concept

If you were asked to make a stem and leaf plot of the number of hours that Brian worked each month, given the data below, would you know or remember how? By the end of this lesson you will!

Make a stem plot of the number of hours that Brian worked each month:

Hours per month: 137, 149, 142, 121, 135, 133, 138, 137, 129, 139, 126, 139

### Watch This



#### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63757>

<http://youtu.be/LEFE1km5ROY> Khan Academy - Reading Box and Whisker Plots

### Guidance

Stem-and-Leaf plots, also known as *stem plots*, are similar to sideways histograms, but are more descriptive because they are built by displaying the specific values of individual data points. Histograms are more visual and can be more striking and/or creative, but stem plots can be more accurate and can make further study of the data more efficient.

Creating a stem plot is not difficult, but, like most things, is easier and quicker with practice. To begin creating your stem plot, start by making two columns and labeling one *leaf* and the other *stem*. You can either create a full table, like this one:

**TABLE 4.57:**

Stem	Leaf
0	0
0	0
0	0

Or you can make one out of simple characters, like this:

Stem	Leaf
0	0
0	0
0	0

Just be sure that it is easy to see which column your values are placed in when the data is entered.

Depending on the size of your numbers, your stem column will either represent the single greatest place value of your data points, or perhaps the two or more greatest place values. The leaf column will represent the smallest place value(s) of the set. If you are listing multiple place values in either column, make a key such as the one below easily visible.

Key: 53,243

53 | 243

If you have two different related sets of data to compare (or *bivariate* data), you can use a *back-to-back* version of the plot. The back-to-back version places the stem in the center of three columns, and the different leaves on each side, as below. Note that ordering the data on the plot requires that you arrange the values so that they increase *from the center out*, rather than just from left to right.

**TABLE 4.58:**

Leaf	Stem	Leaf
3 2 1 0	0	0 1 2 3
3 2 1 0	1	0 1 2 3
3 2 1 0	2	0 1 2 3

Whatever the purpose of your stem plot, always be certain to *label the plot clearly!*

#### Example A

Make a stem plot of the money that Jelani spent this month:

Jelani's purchases: \$59, \$54, \$49, \$68, \$62, \$74, \$57, \$30, \$40, \$80, \$86

**Solution:** Start by creating your table with the heading, stem and leaf columns labeled:

**TABLE 4.59:** Jelani's Purchases for the Month

Stem (tens)	Leaf (ones)

Then begin to insert the data, creating a single row for each tens place:

**TABLE 4.60:** Jelani's Purchases for the Month

Stem (tens)	Leaf (ones)
3	
4	
5	
6	
7	
8	

Finally, fill in the leaves with the ones place values indicated by your data, listing them in increasing numerical order:

**TABLE 4.61:** Jelani's Purchases for the Month

Stem (tens)	Leaf (ones)
3	0
4	0 9
5	4 7 9
6	2 8
7	4
8	0 6

### Example B

Beryl loves basketball, and she tracked the number of points her favorite pro player scored in the first 25 games of his last season. Create a stem plot of the data.

Points per game: 25, 41, 47, 42, 20, 37, 38, 40, 33, 21, 27, 46, 42, 28, 23, 35, 29, 42, 31, 45, 25, 27, 33, 37, 42

**Solution:** Notice that the data points are all between 20 and 50, so we will need only 3 stems in our plot, each of which will have quite a few leaves. Because there are so many points, it might be worth the time to first organize the data in increasing numerical order:

20, 21, 23, 25, 25, 27, 27, 28, 29, 31, 33, 33, 35, 37, 37, 38, 40, 41, 42, 42, 42, 42, 45, 45, 47

Now it is a simple matter to place all of the correct leaves in the proper column:

**TABLE 4.62:** Points Scored per Game

Stem (tens)	Leaf (ones)
2	0 1 3 5 5 7 7 8 9
3	1 3 3 5 7 7 8
4	0 1 2 2 2 5 5 7

### Example C

Create a stem plot to compare the hours that Bruna and Giselle each worked per week during the same three months of last year:

Bruna hours worked: 40, 31, 40, 31, 36, 36, 20, 21, 40, 33, 31, 50, 47, 34

Giselle hours worked: 28, 23, 35, 29, 42, 31, 45, 25, 27, 34, 31, 40, 32, 41

**Solution:** First sort the data in ascending numerical order:

Bruna hours worked: 20, 21, 31, 31, 31, 33, 34, 36, 36, 40, 40, 40, 47, 50

Giselle hours worked: 23, 25, 27, 28, 29, 31, 31, 32, 34, 35, 40, 41, 42, 45

Next create a labeled three column table with the tens column in the center and the ones column for each person on the sides, note that all of the values are between 20 and 50, so we will need only 4 rows. List the values in increasing numerical order from the center out!

**TABLE 4.63:** Hours Worked per Week

Bruna		Giselle
1 0	2	3 5 7 8 9

TABLE 4.63: (continued)

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

**Concept Problem Revisited**

Make a stem plot of the number of hours that Brian worked each month:

Hours per month: 137, 149, 142, 121, 135, 133, 138, 137, 129, 139, 126, 139

First arrange the values in ascending numerical order:

121, 126, 129, 133, 135, 137, 137, 138, 139, 139, 142, 149

Now create your table, listing the first *two* place values in the stem column (since these are all three-digit numbers):

Hours Brian Worked Each Month:

12		1	6	9	
13		3	5	7	7 8 9 9
14		2	9		

Key: 147 = 14 | 7

**Vocabulary**

**Stem Plots** are also known as stem-and-leaf plots because they are a method of arranging data by separating the greatest place value(s) (the stems) from the least place value(s) (the leaves).

A **Stem** in a stem plot is a values or column of values that represent the greatest place value(s) in a set of data.

A **Leaf** in a stem-and-leaf plot is a value or column of values that represent the smallest place values in a set of data.

A **Back-to-Back stem plot** is a modified stem-and-leaf plot with the stem in the center and the leaves on the sides, it is used to compare two different related sets of data (bivariate data).

**Guided Practice**

1. If you asked to make a stem and leaf plot from the following test scores, what would be the first step? 73, 99, 92, 91, 85, 82, 76, 92, 80, 70, 9, 59, 91, 95, 86, 77, 78, 71, 64, 82, 98, 65, 56
2. What is the range of the data? Once the range has been identified, what is going to be the range of the stem and why?
3. Create a Stem-and-Leaf plot for this data.
4. What type of graph would this be if it were turned vertically?
5. In which interval did the least number of students score? In which did the most students score?

**Solutions:**

1. The first step would be to put the scores in numerical order so that they looked like this: 56, 59, 64, 65, 69, 70, 71, 73, 76, 77, 78, 80, 82, 82, 85, 86, 91, 91, 92, 92, 95, 98, 99
2. The range of the data is from 56 to 99. Therefore the stems, identified by the 10's place value, will range from 5 to 9
3. Your plot should look like this:



TABLE 4.64:

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

## 4. A Histogram

5. The first stem: 50 to 59, contains the least number of values, 2 of them. The 70 - 79, and 90 - 99 stems each had 6 scores, for the high.

**Practice**

Make a stem-and-leaf plot from the following data. The data represents the number of times in the last year that each polled student attended a school function.

Number of functions attended: 17, 16, 16, 16, 16, 15, 15, 19, 18, 17, 15, 13, 14, 15, 23, 12, 36, 48, 45, 28, 23, 24, 18, 17, 16, 15, 16, 13, 17, 19, 14, 15, 17, 12

1. Based on the data, how many total students were polled?
2. What interval represents the least number of functions attended by students?
3. What interval represents the greatest number of functions attended by students?

Create stem-and-leaf plots for the following sets of data.

4. Fifteen people were asked how often they bought flavored coffee on their way to work during the last 10 working days. The number of times each person bought flavored coffee was as follows: 5, 7, 9, 9, 3, 5, 1, 0, 0, 4, 3, 7, 2, 9, 8
5. A middle school teacher asked her students how many books they had read over the last year. Their reported answers were as follows: 13, 22, 20, 5, 11, 6, 14, 26, 20, and 13.
6. Joe is training for a swim meet, and keeps track of the number of laps he swims each day. Create a stem-and-leaf plot to organize the data: 21, 22, 23, 18, 28, 27, 24, 24, 28, 29, 24, 20, 27, 28, 21, 40, 21, 9, 27, 23, 28, 27, 25, 27, 19, 31, 30, 24, 30, 28.
7. Redraw the stem and leaf plot, using the above data using 5-unit intervals.
8. Comment on how the data looks different, and explain which chart you feel is a better representation, and why.

The weights to the nearest kilogram of 30 kids were recorded. 59.1, 61.6, 62.2, 61.5, 60.8, 59.9, 60.4, 59.1, 61.0, 60.8, 61.5, 56.2, 61.8, 65.6, 60.3, 58.8, 59.0, 61.3, 62.0, 61.3, 58.5, 60.9, 60.3, 62.6, 60.1, 59.2, 61.8, 61.8, 58.5, 62.1

9. What numbers will be represented by the stem? What numbers will be represented by the leaf?
10. What is the data range? Where will the stems start and finish?
11. Create a stem-and-leaf plot for the data.
12. Which weights would be considered outliers?

You are the football coach & you have two place-kickers. Manny has *made* field-goals of 23 yards, 34 yards, 30 yards, 23 yards, and 28 yards, but has *missed* field goals of 16, 16 and 23 yards. The other, Jose, has *made* field goals of 11 yards, 16 yards, 21 yards, 20 yards, but has *missed* field goals of 26 and 28 yards. There is only 15 seconds left on the play clock.

13. Create Stem-and-Leaf plots to represent the Manny's data.
14. Create Stem-and-Leaf plots to represent the Jose's data.
15. Your team can win the state championship if your kicker makes an 18 yard field-goal. Which kicker would you choose & why?

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.10.

## 4.11 Interpreting Stem-and-Leaf Plots

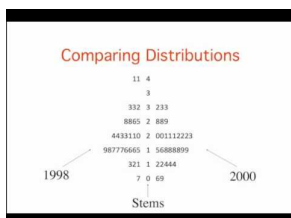
### Objective

Here you will learn and/or practice reading stem plots in order to efficiently extract the relevant information they contain.

### Concept

One of the benefits of stem-and-leaf plots is the clarity with which they display the actual numerical values of data points, while also presenting a clear visual indicator of the *trends* in the set. If you were given a stem plot depicting the money spent each week for a few months, how would you go about finding the parameters of the sample?

### Watch This



### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63759>

<http://youtu.be/zBPRAx0UIOc> Onlinestatbook - Graphing Distributions - Stem and Leaf Displays

### Guidance

Stem plots can be a very efficient means of displaying the data in a set or sample. In fact, stem plots can be thought of as histograms 'drawn' by the data values themselves, rather than by bars representing those values, with *bins* separated by place value rather than otherwise designated intervals. From that point of view, stem plots are a type of histogram that may show more detailed information than standard histograms do, when the separation of bins by place value is appropriate (and it often is not). That means that many of the skills and terminology you have used with histograms applies also to stem plots.

As with histograms, *range* is an important characteristic in a stem plot. The range is the difference between the least and greatest values in the set. It can be important to note when a set appears to have a very large range, especially when there are very limited values in each stem, because it may indicate questionable data.

The shape of a stem plot carries the same general characteristics as a similar shape would if using a histogram:

**Bell-shaped:** An obvious single and central area of the stem plot that has notably more members than the extremes do is referred to as a bell-shaped plot. This shape indicates that most of the values cluster around the median, and quickly become less dense as we move away from the median and toward either of the extremes.

**Uniform:** A consistent width of each leaf suggests data that is not dramatically changed by the input. A very uniform stem plot is not particularly useful for identifying trends in the data, and may suggest a need to increase the number of place values in the stem in order to show more detail in each leaf. For example, a stem plot that appears nearly

uniform with a stem only representing the hundreds place may show significantly more detail if the stem were set to represent the hundreds and tens places both, since this would create many more leaves.

**Skewed:** As with the horizontal skewing of a histogram, stem plots with a obvious skew toward one end or the other tend to indicate an increased number of outliers either lesser than the mode (skewed down - correlating to a left-skew in a histogram) or greater than the mode (skewed up - correlating to a right-skewed histogram). Plots with an upward skew will have a mode that is smaller than either the mean or the median, and a mean that is greater than either the median or the mode. Downward-skewed plots will have a mean lesser than median or mode and a mode greater than either mean or median.

### Example A

Using the stem plot below, make a list of the number of hours that Isolde worked each week:

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

**Solution:** Use the values in the left column to represent the tens place, and the values in the right column to represent the ones, and write each number out as a list:

Isolde's work record (hours per week):

24, 24, 29, 32, 32, 34, 37, 38, 40, 41, 41, 42, 42, 43, 44, 50, 51, 56, 58

### Example B

Describe the type and meaning of the shape of the stem plot, identify the extremes, and state the mode(s) of the data. What does the data describe, as stated by the stem plot?

**TABLE 4.65:** Hathor's Expenses for Jan 2011

<b>1</b>	2 4 5 5 8
<b>2</b>	3 3 6 7 7 9
<b>3</b>	0 0 3 3 3 7 8 8
<b>4</b>	1 4 4 7
<b>5</b>	0 2 3 3

**Solution:** The data is bell-shaped, indicating that the majority of the data is clustered around the median. The median is 33, since there are 13 values above and below that point, and the extremes are 12 and 53 since those are the greatest and least values in the sample. The mode is 33, since there are 3 points with that value and no other value has more than 2 entries.

The plot describes the money spent in the month of January 2011, by someone named Hathor. Hathor bought mainly items less than \$40, though there were a few purchases above the \$40 mark, none were greater than \$53.

### Example C

Compare the weekly incomes of Sabah and Anaelle according to the stem plot below.

- What things stand out based on the shapes of the two plots, if any?
- What is the place value of the stem?
- What does the blank on Anaelle's side next to the "1" indicate?
- Who earned the most money overall?
- Who had the most consistent paychecks?

- f. Who had the single greatest and who the single least paycheck?
- g. Did they both record the same number of paychecks?

**TABLE 4.66:** Week Income Comparison (USD)

Sabah		Anaelle
99 97 88	0	89 95
79 23	1	
93 68 42	2	09 17 22 58 77
97 58 12 00	3	01 29 33 51 67 82

Key: 147 = 1 | 47

**Solution:** Let's take each of the questions individually:

- a. Both plots are slightly up-skewed, indicating that the lower values are likely less indicative of the most common incomes. Sabah's plot is otherwise pretty uniform, whereas Anaelle's has a notably greater frequency of values in the \$200 to \$400 range.
- b. The stem represents the hundreds place.
- c. Anaelle did not record any paychecks between \$100 and \$200.
- d. Without even totaling the columns, we can see that Anaelle earned more, as her column is heavily weighted towards the greater values, and she has also recorded an additional check.
- e. Anaelle, as indicated by the greater number of checks clustered in the \$200 - \$400 range and few outliers.
- f. Sabah had the smallest paycheck at \$88, and also the greatest at \$397.
- g. No, Anaelle recorded 13, Sabah only 12.

### Concept Problem Revisited

If you were given a stem plot depicting the money spent each week for a few months, how would you go about finding the various important parameters of the sample?

First look over the general shape of the plot to get an idea of the trend(s) of the data. Then identify the maximum and minimum extremes and the median and mode. Ask yourself if there are any particularly extreme outliers, or if the data is spread evenly, if there is there an obvious mode, if the median and the mode are close together, and if there are any other standout values. You can quickly get a good picture of the data by understanding how a stem plot works and applying your knowledge, but the data is useless if you don't know how to read the data or what questions to ask.

### Vocabulary

**Trends** in data sets or samples are indicators found by reviewing the data from a general or overall standpoint. Values such as the 5 number summary can help to identify trends such as increasing or decreasing values over time or fluctuations in values caused by a particular factor.

**Bins** are the intervals or categories in histograms, the name comes from the idea that you review the data and toss each point into a corresponding 'bin' based on which range or category it fits.

The **range** of a set of data is the difference in value between the least (the minimum extreme) and greatest (the maximum extreme) values in the set. Investigating the range of a set can help to determine how widely spaced the data points are.

**Guided Practice**

The stem and leaf plot below shows the grade point averages of 18 students.

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

1. What is the range of the data in the stem and leaf plot?
2. How many students have a grade of 2 or more?
3. What is the mode of the grades?
4. What is the median of the grades?

**Solutions:**

1. Range = maximum value – minimum value =  $4.0 - 0.8 = 3.2$
2.  $7 + 4 + 1 = 12$  students
3. Two modes: 1.4 and 2.5
4. There are 18 data values and they are already ordered in the stem and leaf diagram.

$$\text{Median} = \frac{(\text{the 9th value} + \text{the 10th value})}{2} = \frac{(2.5 + 2.5)}{2} = 2.5.$$

**Practice****True or False**

1. Some sets of data do not have a mean.
2. There is always a mode for each set of data
3. The median is always a number in the data set.

Based on the following stem-and-leaf plot of student heights in inches, how tall is tallest person in class?

4.

**TABLE 4.67:**

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

Taylor and some friends went crawdad hunting. The number of crawdads captured have been recorded below.

**TABLE 4.68:**

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

5. How many people went with Taylor?
6. What were the total number of crawdads captured?

The following stem- and-leaf plot shows a range of 10 numbers.

**TABLE 4.69:**

87	5
88	1 4 5
89	1 7 7
90	6 9

7. What is the highest number?
8. What is the median value of the data set?

The back to back stem and leaf plot below shows the exam grades (out of 100) of two different class periods. The digit in the stem represents the tens and the digit in the leaf represents the ones.

**TABLE 4.70:**

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

9. How many students scored higher than 60 in section 1?
10. How many students scored higher than 60 in section 2?
11. What are the minimum and maximum scores in section 1?
12. What are the minimum and maximum scores in section 2?
13. Without counting, which section has more students scoring 80 or more?
14. Without counting, which section has more students scoring 50 or less?
15. Describe the shape of the data set. Is it symmetric? Are there extreme values in the low or high numbers?

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.11.

## 4.12 Creating Scatter Plots and Line Graphs

### Objective

Here you will learn how to take raw or organized *bivariate* data and present it in a visual format with a scatter plot or line plot.

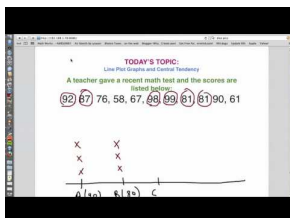
### Concept

Scott's teacher was reviewing the research that Scott had conducted regarding the best car stereo systems to buy for college students on a limited budget. Scott had two long columns of numbers indicating the comparison between sound quality (which Scott had summarized with a 10-point scale for each stereo), and cost rounded to the nearest dollar before tax.

The teacher commended Scott on the detailed research, but pointed out that the list of numbers was kind of hard to make sense out of. He suggested that Scott plot the values on a scatter plot or line graph to see if there was a 'sweet spot' indicating the best compromise between quality and cost.

How should Scott decide which type of plot is best for his purpose? How would he go about taking the data from columnar form and converting it into the data visualization he decides to use?

### Watch This



### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63762>

<http://youtu.be/TZGOIeKp0fc> MathMasterD - Line Plot Graph Tutorial Video

### Guidance

Line plots, followed closely by scatter plots, are by far the most common method of displaying *bivariate data*. By assigning one variable to each axis and plotting points by both horizontal and vertical location simultaneously, you can quickly and easily show the degree to which one set of data is influenced (or not influenced) by another.

There are two general types of bivariate data sets that are graphed on a line or scatter plot: observed (or experimental) data and calculated (or predicted) data.

- **Calculated Data:** To create a line or scatter plot of calculated data, you must first identify your two variables as either *dependent* or *independent*. A dependent (or input) variable may also be referred to as the *explanatory* variable, and has values that are assigned to it. An independent (or output) variable may also be called the *response* variable, and has values that are the result of computations performed on the input variable. By



convention, the independent variable is plotted on the horizontal, and the dependent variable is plotted on the vertical.

- **Observed Data:** The most common reason to graph two sets of data on the same graph is to evaluate the level of *statistical correlation*. By plotting the two sets of data on separate axes of the same graph, we can see a visual representation of possible related changes in values between the two sets. As with calculated data, you should plot the values of the variable that you expect is the explanatory variable on the horizontal and the expected response variable on the vertical.

When graphing observed data, you do not always know which value is the input and which the output, or even if the two values are indeed dependent at all! In later lessons, we will return to this concept to learn a number of methods to evaluate data and determine the degree of correlation between multiple value sets. For now, place the variable you think is most reasonably the input on the horizontal.

The first and most important step is to organize your data so that it is easy to see how a given input value relates to a given output value. By convention this is done with a 'T' chart or a two-column graph, with the input value on the left and the output value on the right, or vertically with the input on the top and output on the bottom.

Once you have the table constructed, start with the first pair of values and move across your horizontal axis to the first input value and up the vertical axis to the associated output value. Continue the process until all of your points have been graphed.

Once all of your points have been plotted, if you are creating a scatter plot, you're done! If you are creating a line plot, start at your minimum input value and connect the points as you move to the right on the input axis.

### Example A

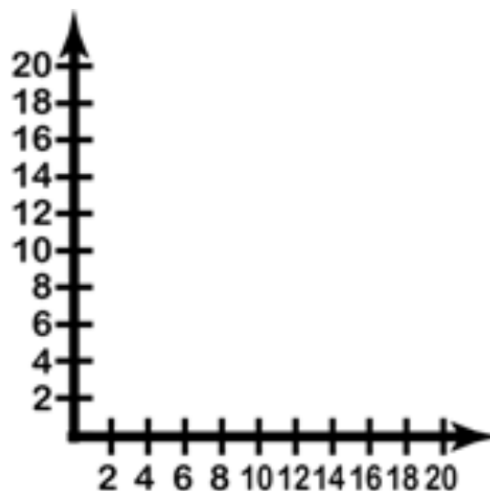
Construct a scatter plot from the given values.

**TABLE 4.71:**

<b>Input</b>	<b>1</b>	<b>3</b>	<b>5</b>	<b>7</b>	<b>9</b>	<b>11</b>	<b>13</b>	<b>15</b>	<b>17</b>
<b>Output</b>	2	4	6	8	10	12	14	16	18

**Solution:** The data here is already organized into associated input and output values, so you simply need to create a graph with a horizontal and vertical axis on which to plot the points.

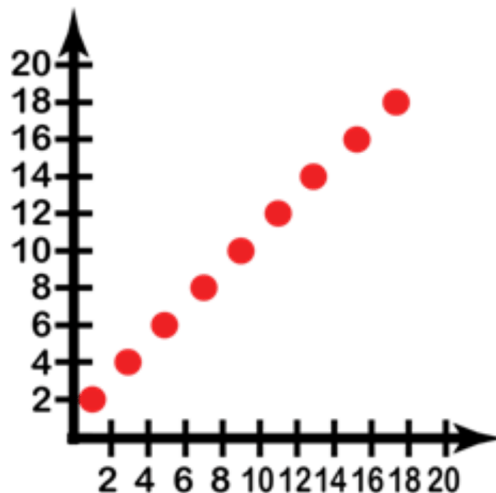
Notice that I have only created the positive values here, since the table of values was all positive.



Now we just plot the points from the table, starting with the first vertical pair: Input = 1, Output = 2. Incidentally, when describing a single point of bivariate data, the conventional method of writing it is in the form (input, output)

or  $(x,y)$ . So our first point would be  $(1, 2)$ , the second would be  $(3, 4)$  and so on.

Now we fill in the values on the graph, starting with  $(1, 2)$ . Beginning at the lower-left corner, which represents  $(0, 0)$ , move 1 point to the right and 2 points up. The second point is 3 points to the right and 4 points up. Continue until all 10 points are graphed. Since the question asks specifically for a scatter plot, once the individual points are plotted, we are done.



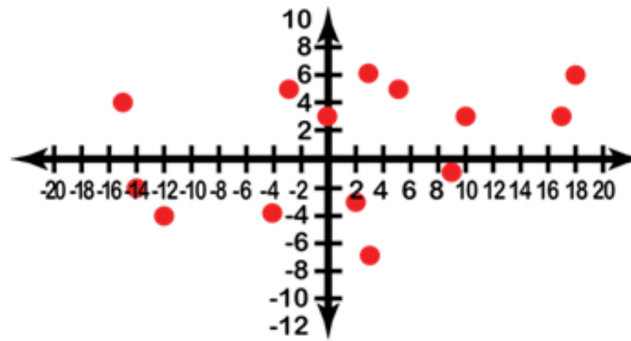
### Example B

Romane loves jellybeans, and she eats an average of 20 each day. Worried about her weight, she decides to see if there is an obvious correlation between the number of jellybeans she eats and her weight. If she records the data below, which variable would be the input and which the output? Create a line plot from the data.

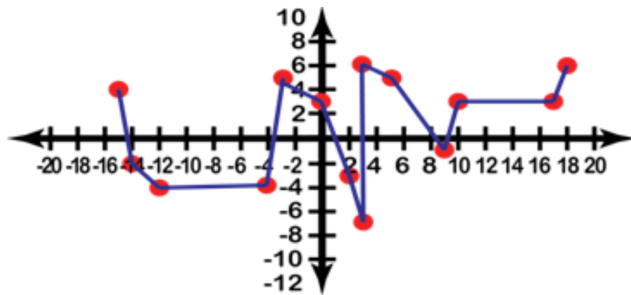
TABLE 4.72:

Increase or Decrease # of beans	Increase or Decrease in Weight
+2	-3
-3	+5
-7	+3
+5	+5
-12	-4
+18	+6
-14	-2
-15	+4
+17	+3
+0	+3
-4	-4
+3	+6
+9	-1
+10	+3

**Solution:** Since Romane can control the number of jellybeans she eats, that would be the input variable, and the increase or decrease in weight would be the output. If we create an  $(x,y)$  graph and plot the points, the result looks like this (note that this time the graph shows negative and positive values!):



Finally we connect the points from left to right, since the question specified a line graph:



### Example C

Does more sleep consistently improve math grades?

Organize the data below by creating a 'T' table for the  $x$  and  $y$  values, then graph the data as either a scatter plot or line graph, whichever is most appropriate.

**TABLE 4.73:**

Math Homework Score (out of 20 points)		Hours of Sleep (the night before)	
Day 1	11	Day 1	7
Day 2	19	Day 2	8
Day 3	9.5	Day 3	6.5
Day 4	11	Day 4	7
Day 5	15	Day 5	5
Day 6	6	Day 6	3
Day 7	11.5	Day 7	8.5
Day 8	18.5	Day 8	7.5
Day 9	14	Day 9	6
Day 10	18	Day 10	8
Day 11	15.5	Day 11	5.5
Day 12	15.5	Day 12	4.5
Day 13	12	Day 13	9
Day 14	15.5	Day 14	5.5

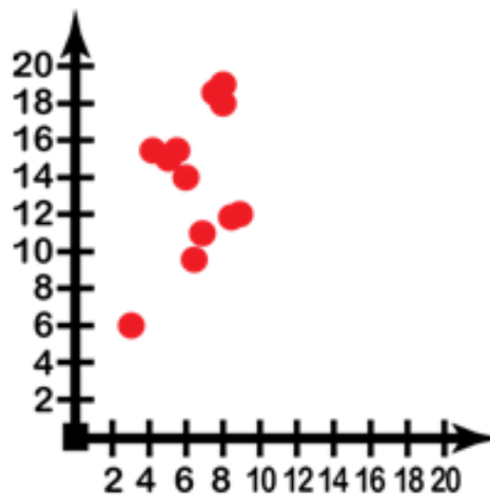
**Solution:** The data as given is organized by day in each table. In order to answer the question: “Does more sleep consistently improve math grades?” We need to correlate the data from each night’s sleep with the next day’s grades.

In this case, *which* day it is does not matter as much as the hours of sleep the night before, so we can pull the 'Day' column out of each table (being careful not to change the order of the values!) and make a new table with only the correlating scores and hours of sleep. This gives us:

TABLE 4.74:

Hours of Sleep	Math Score
7	11
8	19
6.5	9.5
7	11
5	15
3	6
8.5	11.5
7.5	18.5
6	14
8	18
5.5	15.5
4.5	15.5
9	12
5.5	15.5

Since the question is asking about the correlation between sleep the night before and grade the next day, the sleep becomes the *input variable* (or *independent variable*) and score becomes the *output variable* (*dependent variable*). Plotting the points on an  $(x,y)$  graph yields:



Given the significant scattering of points as we move left to right, it is appropriate to maintain the scatter plot layout.

In later lessons we will discuss *linear regression*, the process of identifying a *line of best fit*. A *line of fit* is a line drawn through a scatter plot that indicates a trend that the data follows, and the line of best fit is the mathematically derived most accurate indicator of that trend.

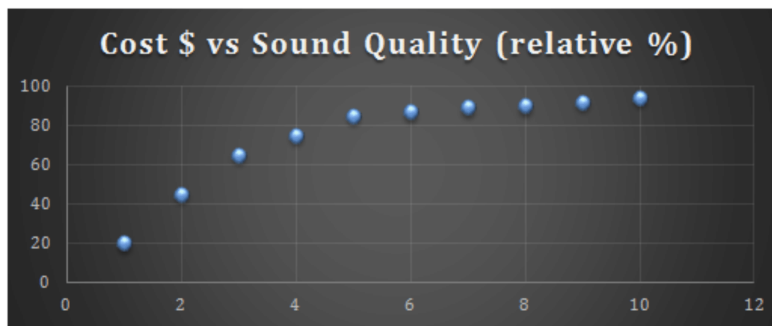
### Concept Problem Revisited

*Scott's teacher was reviewing the research that Scott had conducted regarding the best car stereo systems to buy for college students on a limited budget. Scott had two long columns of numbers indicating the comparison between*

sound quality (which Scott had summarized with a 10-point scale for each stereo), and cost rounded to the nearest dollar before tax.

The teacher commended Scott on the detailed research, but pointed out that the list of numbers was kind of hard to make sense out of. He suggested that Scott plot the values on a scatter plot or line graph to see if there was a 'sweet spot' indicating the best compromise between quality and cost. How should Scott decide which type of plot is best for his purpose? How would he go about taking the data from columnar form and converting it into the data visualization he decides to use?

Scott should view cost as the independent variable (the input) and sound quality as the dependent variable (the output). By creating an  $(x,y)$  graph and plotting each of the corresponding cost/quality values, he will generate a much clearer comparison for the audience of his report. If he notes a particularly clear correlation (all or most of the points in a line) between increasing cost and improved quality, he may wish to indicate a line of fit to further illustrate the trade-off. A graph of his data might look something like this:



## Vocabulary

**Bivariate data** is data composed of two changing sets of values. Distance travelled over differing time intervals, or weight compared to height, or income related to age are all examples of bivariate data.

**Dependent variables**, also called **output variables** or **response variables** and commonly represented by 'y', have values that depend on the value of another variable.

**Independent variables**, also called **input variables** or **explanatory variables** and commonly represented by 'x', have values that are not determined by another variable.

A **line of fit** is a straight or continuously curved line representing the trend of changes in the comparison of two data sets (or one set of bivariate data). **The line of best fit** is the mathematically calculated most accurate depiction of the trend(s). Note that a line of fit need not be straight, or even continuous.

**Linear regression** is the process of identifying a line of fit or the line of best fit for a given function.

## Guided Practice

1. Construct a scatter plot to represent the data from the chart below indicating the number of birds killed by planes each year.

**TABLE 4.75:**

YEAR	PLANES REGISTERED	BIRDS KILLED
1978	6	13
1979	4	12
1980	7	14

**TABLE 4.75:** (continued)

1981	3	11
1982	7	14
1983	6	13
1984	3	12
1985	4	11
1986	1	9
1987	4	12

2. Construct a line graph to illustrate the data.

**TABLE 4.76:**

Height Change	4	2	8	13	16
Weight Change	3	3	3	4	3

3. Mike decided to see if the teenage drivers in his city were truly more likely to get into accidents, and he collected the data below. Graph the data appropriately for his study.

**TABLE 4.77:**

<b>Year</b>	<b>Teen Drivers Registered</b>	<b>Auto Accidents</b>
1980	90	80
1981	70	90
1982	60	90
1983	20	100
1984	30	90
1985	100	90
1986	70	100
1987	60	100
1988	70	100
1989	90	90

4. How do you determine which values to graph on the vertical axis and which on the horizontal?

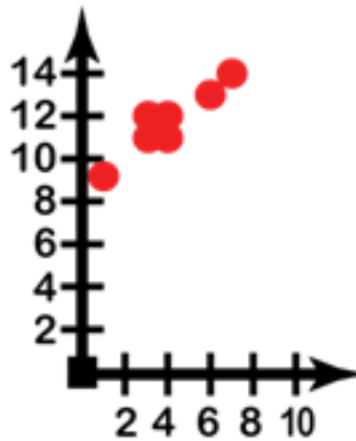
5. Given the data below, which variable represents the explanatory variable? What is the related term for the other variable? How do you know which is which?

**TABLE 4.78:**

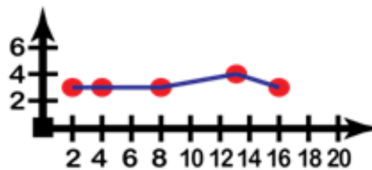
Sneezes	8	6	5	2	1	4	7	12
Tissues	18	13	7	5	1	9	19	43

Solutions:

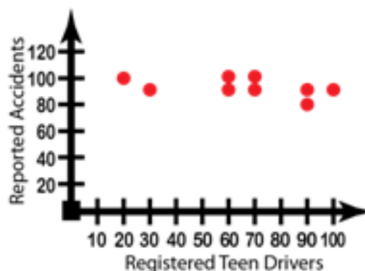
1. The number of planes registered is the input, and the number of birds killed is the output.



2. Change in height is the input, and change in weight the output. Connect the points to create a line graph instead of a scatter plot.



3. The number of teens registering to drive is the input, and the number of reported accidents the output. Note that the very little vertical change despite a significant horizontal change would indicate virtually no correlation between the two values.



4. The input value or the cause is the independent variable, and the output or the effect is the dependent variable. The independent variable goes on the horizontal and the dependent on the vertical.

5. Sneezes are the input variable, graphed on the horizontal 'X' axis, since sneezing is the cause related to the effect of using a tissue.

### Practice

1. Create a scatter plot of the data shown below. Describe the relationship that exists within the data.

**TABLE 4.79:** Cost associated with raising a child

Child's Age	3	6	9	12	15
Annual Cost	11,800	12,800	13,700	16,000	17,800

2. Create a scatter plot from the data in the table below.

**TABLE 4.80:**

X	-3	-3	-2	-1	0	0	0	1	1	1	2	3
Y	3	4	3	2	1	2	0	0	-1	-1	-3	-4

3. Draw a reasonable line of fit.

4. What is the equation of the line of fit?

5. The data below shows the number of hours spent studying for a history quiz. Draw a scatter plot.

**TABLE 4.81:**

Study in Hours	4	3	6	2	1	5	4
Grade in Percent	85	78	93	71	61	91	76

6. Draw a reasonable line of fit.

7. What is the equation of the line of best fit?

8. Predict the grade for a student who studied 7 hours.

9. Could the line go on forever? Why or why not?

10. The table below shows the number of reported food poisoning cases at a local hospital. Create a scatter plot for the data.

**TABLE 4.82:** Reported Food Poisoning Gases

Year	2005	2006	2007	2008	2009
Cases	38	26	19	15	17

11. What relationship, if any exists in the data?

12. Draw a line of fit. Write the slope intercept form of an equation for the line of fit.

13. The table shows the average and maximum lifespan of animals that are kept in captivity. Create a scatter plot to represent the data.

**TABLE 4.83:** Lifespan in Years

Average	13	26	16	9	36	41	42	21
Maximum	48	51	41	21	71	78	62	55

14. Draw a reasonable line of fit, and write the slope intercept form of the equation.

15. Predict the maximum lifespan for an animal with an average age of 33 years.

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.12.



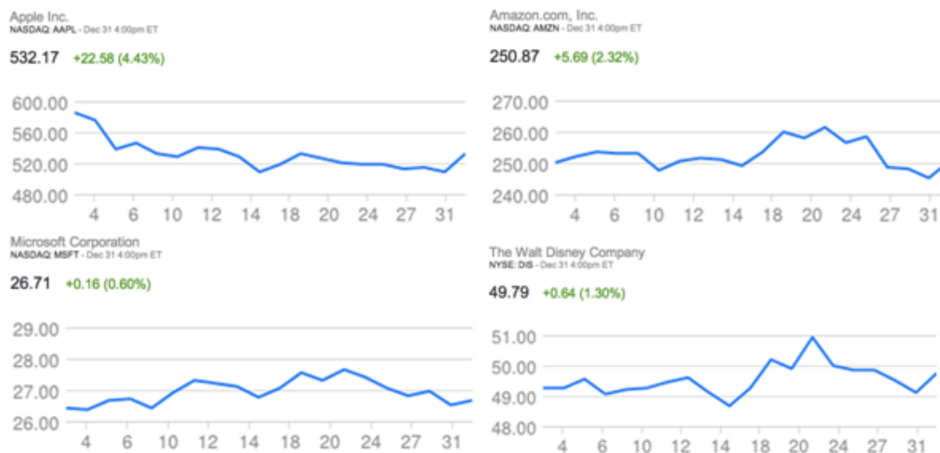
## 4.13 Interpreting Scatter Plots and Line Graphs

### Objective

Here you will practice recognizing and using some of the primary information available from a scatter plot or line graph.

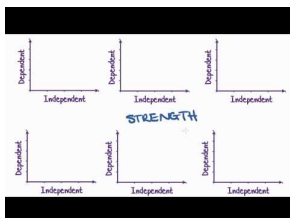
### Concept

Steve is playing a stock market simulation game in his social studies class. He has chosen to invest in Apple Inc., Amazon.Com Inc., Walt Disney Company and Microsoft. He originally bought all four stocks in the 10th day of the month. Now he needs to choose one of them to sell. Based on the recent performance of each of them as shown in the line graphs below, which would you recommend he choose and why?



Seem a bit daunting? It certainly can be! We will return to this question at the end of the lesson to review the situation.

### Watch This



### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63798>

[http://youtu.be/PE\\_BpXTyKCE](http://youtu.be/PE_BpXTyKCE) vcefurthermaths - Maths Tutorial: Interpreting Scatterplots (statistics)

## Guidance

The primary use for scatter plots and line graphs is to demonstrate or evaluate the correlation between two variables. Though the two are similar in many ways, there are distinct differences, and specific situations in which one is appropriate and the other is not.

- A **scatter plot** is generally used when displaying data from two variables that may or may not be directly related, and when neither of the variables is under the direct control of the researcher. The primary function of a scatter plot is to visualize the strength of correlation between the two plotted variables. The number of sunburned swimmers at the local pool each day for a month would be an example of a data set that would best be displayed as a scatter plot, since neither the weather nor the number of swimmers present is under the control of the researcher.
- A **line graph** is appropriate when comparing two variables that are believed to be related, and when one of the variables is under the direct control of the researcher. The primary use of a line graph is to determine the **trend** between the two graphed variables. The mileage of a particular car compared to speed of travel would be a good example, since the mileage is certainly correlated to the speed and the speed can be directly controlled by the researcher.

In later lessons we will discuss methods of quantifying the level of correlation between two variables and calculating a line of best fit, but for now we will focus on identifying specific examples of weak or strong correlation and identifying different types of trends.

- **SCATTER PLOTS:**

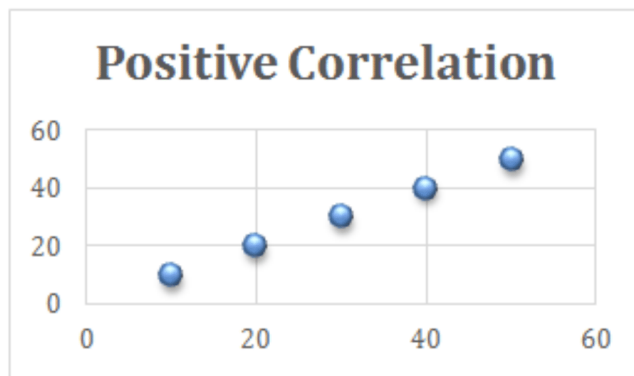
- Two variables with a **strong correlation** will appear as a number of points occurring in a clear and recognizable linear pattern. The line does not need to be straight, but it should be consistent and not exactly horizontal or vertical.
- Two variables with a **weak correlation** will appear as a much more scattered field of points, with only a little indication of points falling into a line of any sort.

- **LINE GRAPHS:**

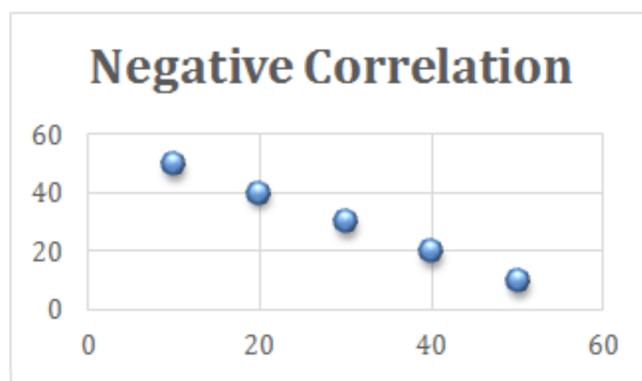
- A **linear relationship** appears as a straight line either rising or falling as the independent variable values increase. If the line rises to the right, it indicates a **direct relationship**. If the line falls to the right, it indicates an **inverse relationship**.
- A **non-linear relationship** may take the form of any number of curved lines, and may indicate a squared relationship (dependent variable is the square of the independent), a square root relationship (dependent variable is the square root of the independent), an inverse square (dependent variable is one divided by the square of the independent), or many other possibilities.

- **BOTH:**

- A **positive correlation** appears as a recognizable line with a positive **slope**. A line has a positive slope when an increase in the independent variable is accompanied by an increase in the dependent variable (the line rises as you move to the right).

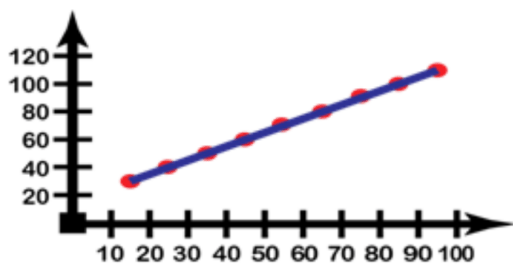


- A **negative correlation** appears as a recognizable line with a negative slope. As the independent variable increases, the dependent variable decreases (the line falls as you move to the right).



#### Example A

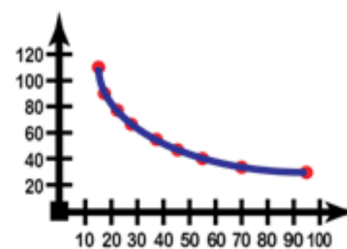
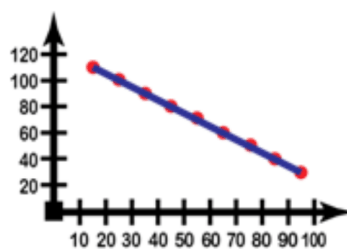
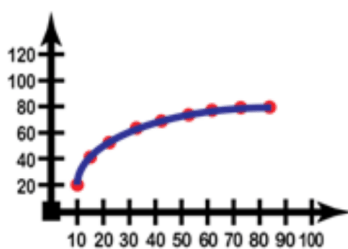
What type of relationship is indicated by the line graph below?



**Solution:** The line is straight, indicating a linear relationship. It rises from left to right, meaning that the dependent variable increases as the independent variable increases, indicating a positive correlation.

#### Example B

Which image shows a non-linear graph with a negative correlation?

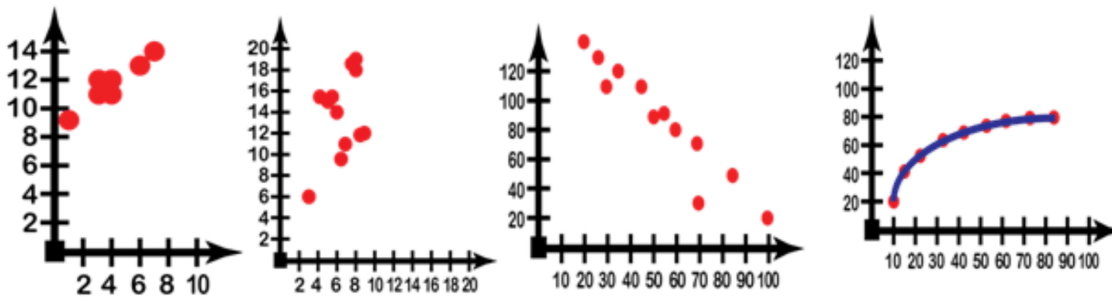


Solution:

- The first image is of a curved line that rises from left to right, this is a non-linear positive correlation
- The second image is a straight line that falls from left to right, this is a linear negative correlation
- The third image is a curved line that falls from left to right, this is a nonlinear negative correlation and is the correct image as described by the question.

### Example C

1. Which graph(s) indicate(s) a weak correlation?
2. Which one(s) indicate(s) a strong correlation?
3. Which graph(s) indicate positive correlation(s)?



Solution:

- Only graph 2 indicates a weak correlation, since it is the only one with points that are not clearly arranged in a linear fashion.
- Graphs 1, 3, and 4 all indicate strong correlations, as evidenced by the high percentage of points obviously organized in a line. Graph 4 is obviously a very strong correlation as a clear non-horizontal or vertical line connects all of the points.
- Graphs 1, 2, and 4 are all positive correlations as all three rise from left to right. Another way to put it is that those three graphs have a positive slope (though graph 4 does not have a consistent slope, anywhere on the curve the slope is estimated it would still be positive).

### Concept Problem Revisited

Which stock(s) should Steve sell if he needs to make a profit right away?

By looking at the lines on each of the four graphs, we can see that it is important to note that Steve purchased the stocks on the 10<sup>th</sup>, since only the Walt Disney CO and Amazon are currently valued more highly than they were on the 10<sup>th</sup>. Both Apple and Microsoft are going up in value now, at the end of the month, but neither has made it back up to where they were on the 10<sup>th</sup>.

If Steve wants to make a profit right now, he should sell Walt Disney or Amazon or both.

### Vocabulary

A **trend** is an estimation of the tendency of data points to move in a certain direction. A **trend line**, also known as a **line of fit**, is a line drawn on a graph to indicate how the data points generally increase or decrease.

A **strong correlation** means that the values of the output variable are strongly affected by the values of the input variable. A strong correlation is indicated on a graph by a large percentage of data points lying in an apparent line, either straight or curved.

A **linear relationship** means that the output values are a simple multiple of the input variable, and appears as a straight line when graphed.

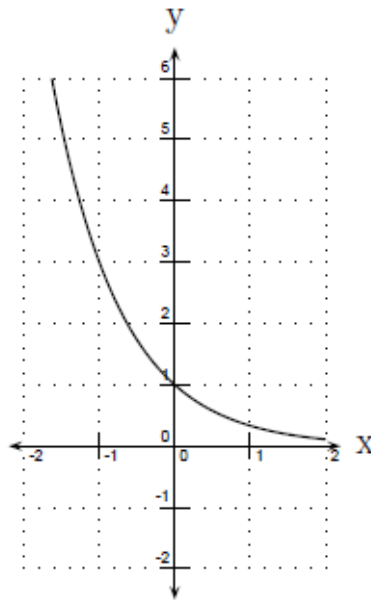
A **non-linear relationship** appears as a curved line on a graph. It indicates an output variable that is a power, a root, or other more complex multiple of the input.

A **direct relationship** means that the variables increase and decrease together, resulting in a **positive correlation** and a line of fit that rises from left to right, whereas an **inverse relationship** is a **negative correlation**, meaning that the output decreases as the input increases, and vice versa.

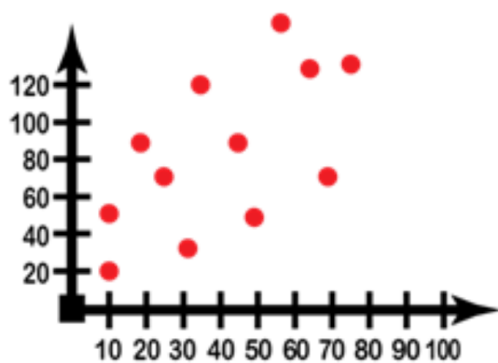
A **slope** is a description of the rate at which the output variable increases or decreases compared to the input variable. This is referred to as the slope because the rate of increase or decrease affects the angle of the line on a graph.

### Guided Practice

1. Describe the relationship indicated by the graph:

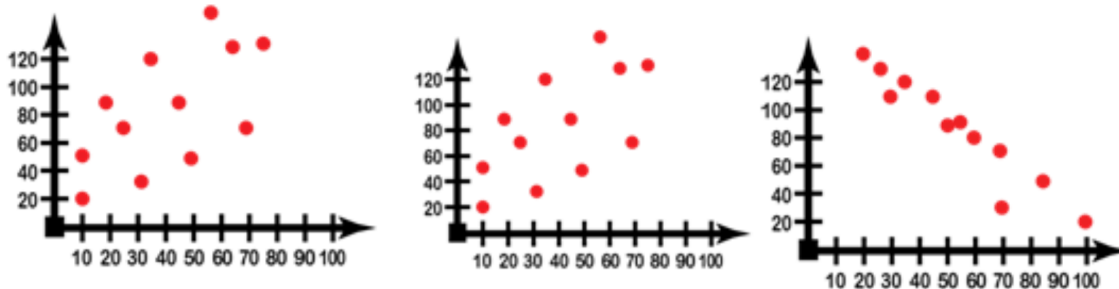


2. Describe the relationship indicated by the graph:

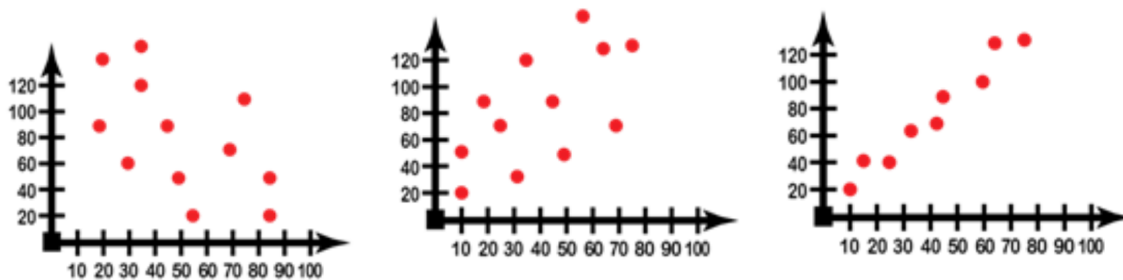


3. Describe the graph that would result from a strongly correlated positive non-linear relationship. Give an example of a function that could result in such a graph.

4. Which scatter plot below indicates the most strongly correlated variables?



5. Which plot below indicates a weakly correlated positive linear relationship?

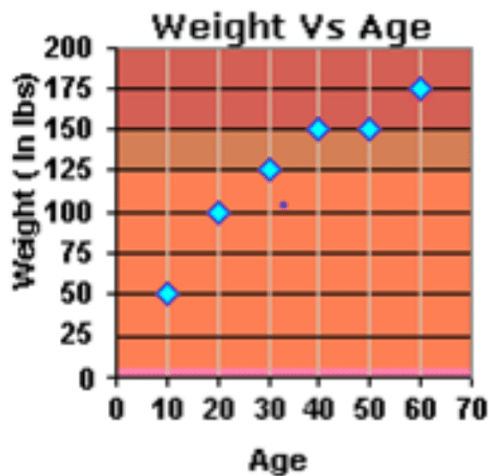


Solutions:

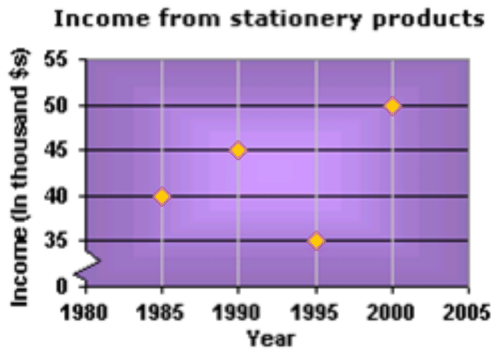
1. This is a very strongly correlated (all the points connected by a line), negative (the line falls from left to right), nonlinear (not a straight line) relationship.
2. This is a weakly correlated (significant scattering of the points), positive (points generally increase in value from left to right), linear (a straight line of fit could be drawn) relationship.
3. A strongly correlated positive non-linear relationship would appear as a well-defined curve of points rising from left to right.
4. The center plot is the most strongly correlated, evidenced by the much cleaner line formed by the data points. Incidentally, this is a negative linear relationship.
5. The left hand plot is weakly correlated, but negative. The right hand plot is positive, but strongly correlated. The center plot is weakly correlated and positive, so it is the one matching the question definition.

**Practice**

1. What sort of trend is shown in the scatter plot below?



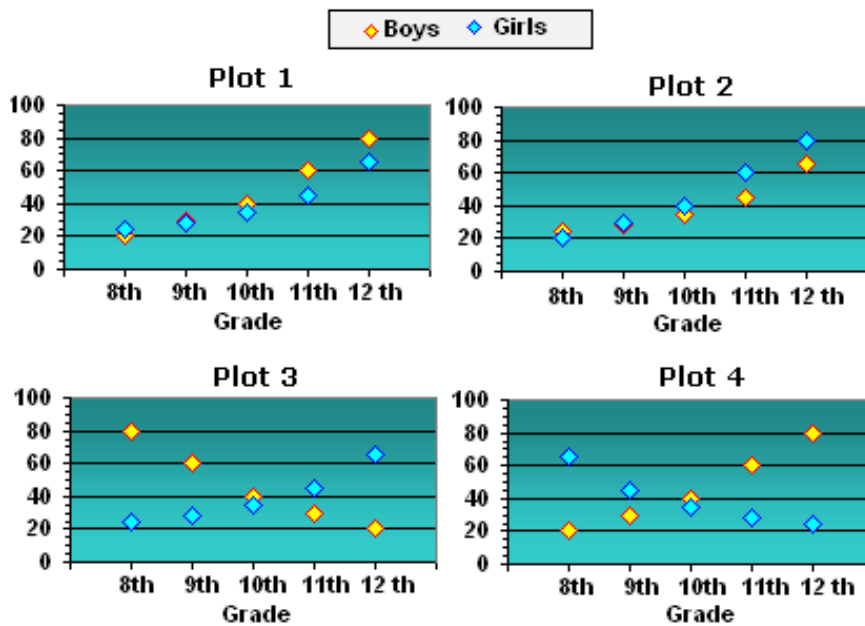
2. A door to door vacuum cleaner sales man plots a scatter diagram of how much he has earned over the years. In which year was his income the highest?



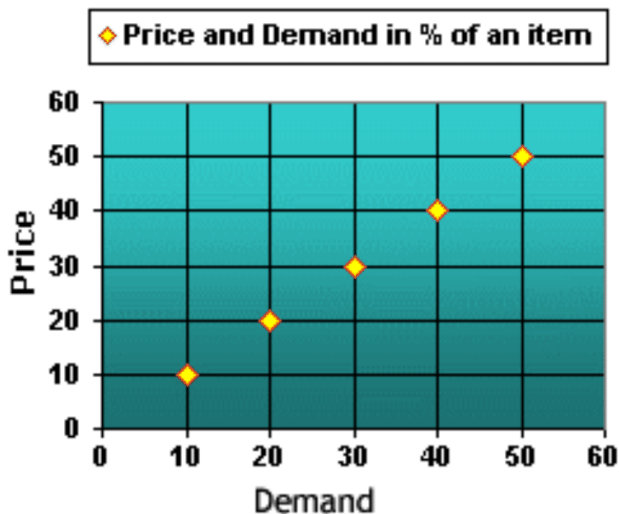
3. The number of children in two different day care centers, and the types of lunch they eat is represented in the table below. Pick the appropriate scatter plot for the data.

TABLE 4.84:

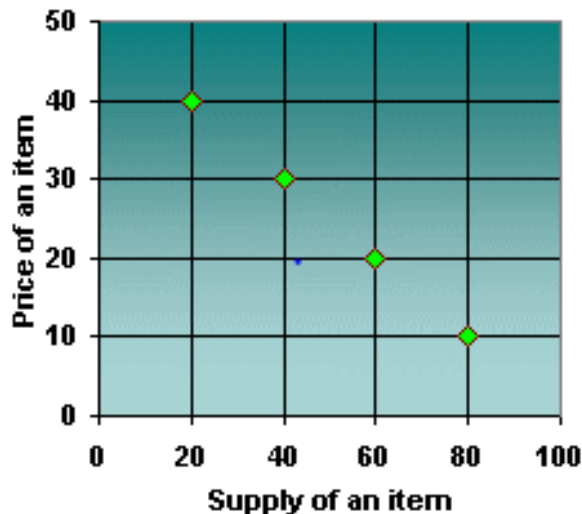
Lunch Served	Center 1 (Yellow)	Center 2 (Blue)
Hamburger	20	25
Mac and Cheese	30	28
Pizza	40	35
Tuna Salad	60	45
Burritos	80	65



4. The plot shown gives the relationship between the demand and price of a trendy consumer good. What trend does the plot follow?



5. The plot represents the relationship between the price and supply of an item. What type of trend does the graph illustrate?



6. Katie recorded the following data relating to how long it took to fill up a horse trough. She measured the depth every two minutes after she began filling it, until it was full. Which scatter plot accurately represents the data?

**TABLE 4.85:**

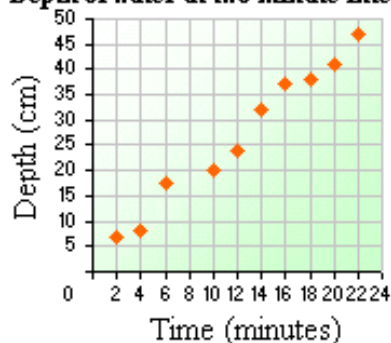
Time (in minutes)	Dept (in inches)
2	7
4	8
6	13
8	19
10	20
12	24
14	32
16	37
18	38
20	41



**TABLE 4.85:** (continued)

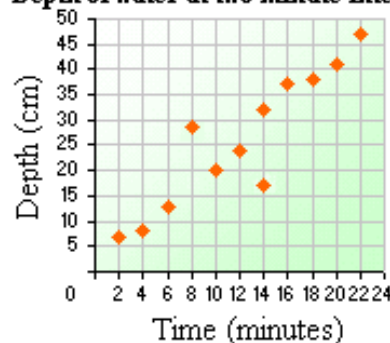
22	47
----	----

**Depth of water at two-minute intervals**



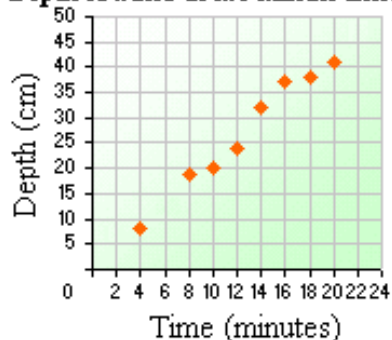
**Graph 1**

**Depth of water at two-minute intervals**



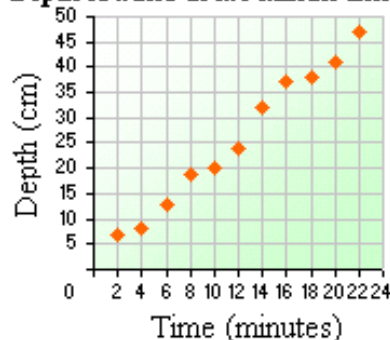
**Graph 2**

**Depth of water at two-minute intervals**



**Graph 3**

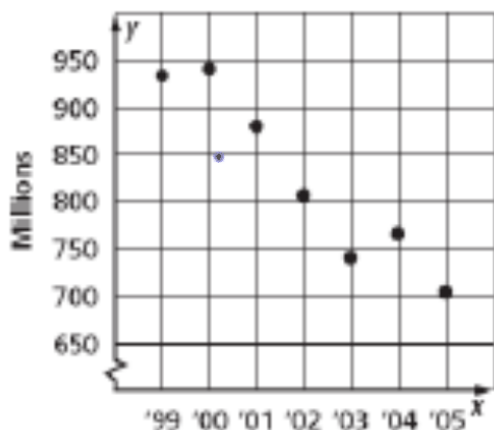
**Depth of water at two-minute intervals**



**Graph 4**

7. The Scatter Plot below question 8 shows the number of DVD's sold (in millions) from 2001-2007. Based on the data, about how many DVD's will be sold in 2009?

8. What sort of trend is shown in the scatter plot?



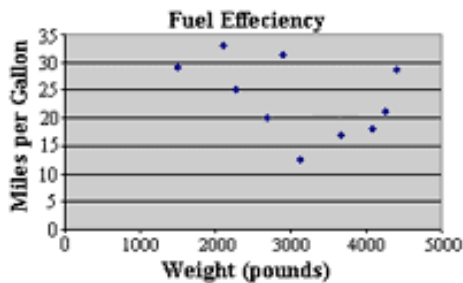
9. The table below shows a relationship between the weight of a car and its average gas mileage. Which plot best represents the data?

**TABLE 4.86:**

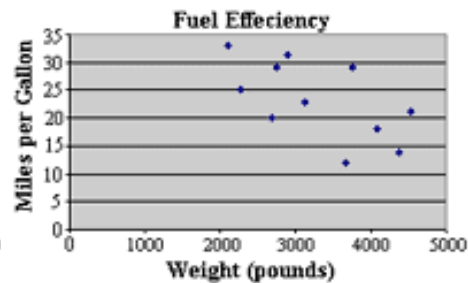
Type of Car	Weight	MPG
-------------	--------	-----

**TABLE 4.86:** (continued)

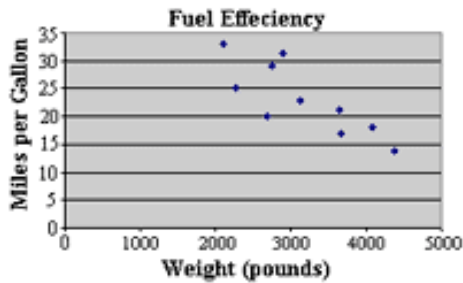
1	3750	29
2	4125	23
3	3100	33
4	5082	18
5	3690	20
6	4640	21
7	5380	14
8	3241	25
9	3895	31
10	4669	17



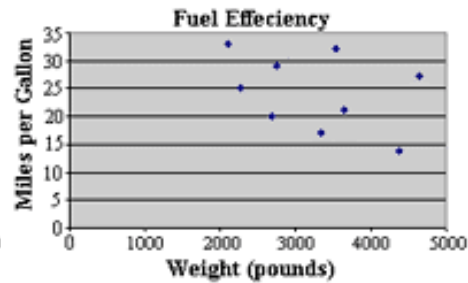
**Graph 1**



**Graph 2**

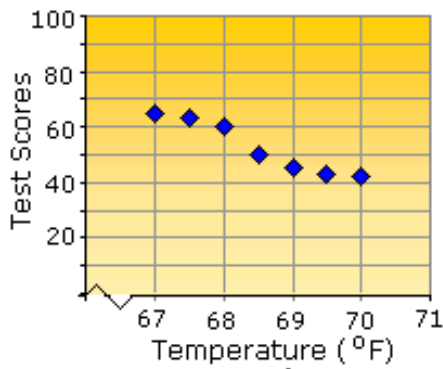


**Graph 3**

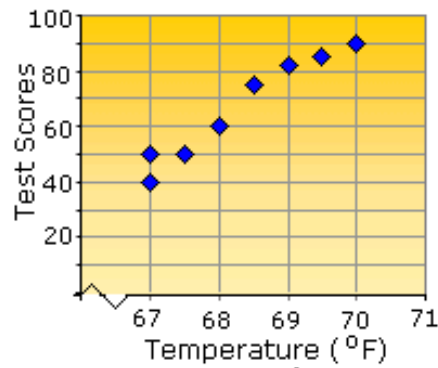


**Graph 4**

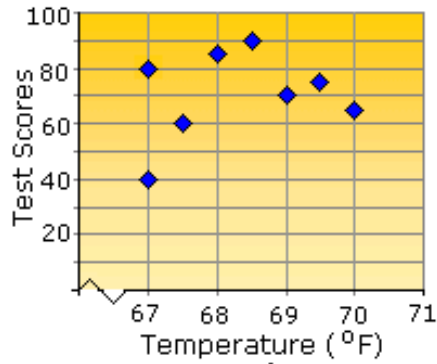
10. Which scatter plot shows no relationship between test scores received by Greg, and the temperature that the classroom was at while taking the test? Why?



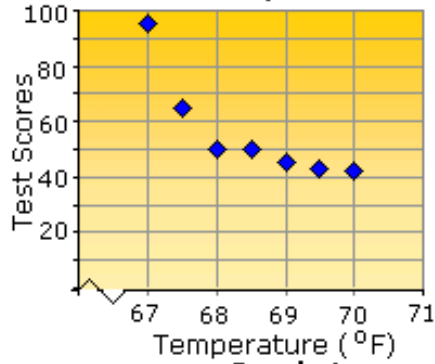
**Graph 1**



**Graph 2**

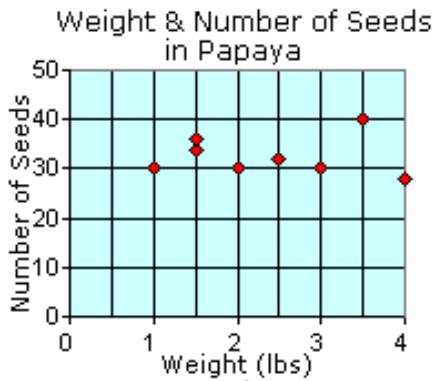


**Graph 3**

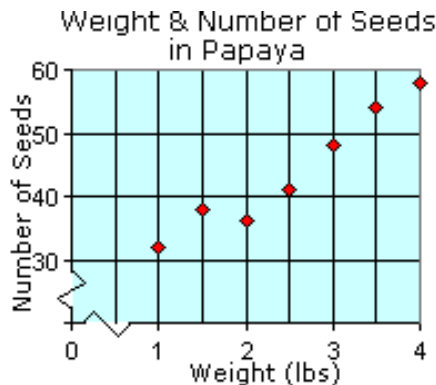


**Graph 4**

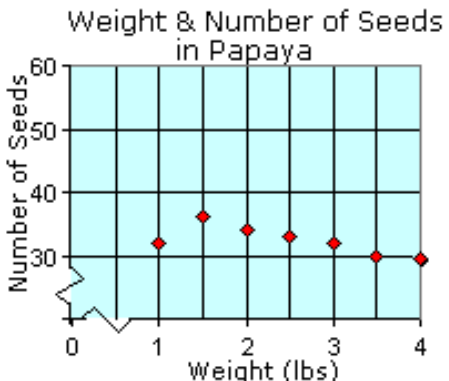
11. Which scatter plot shows a positive relationship between the weight of a mango, and the number of seeds it contains?



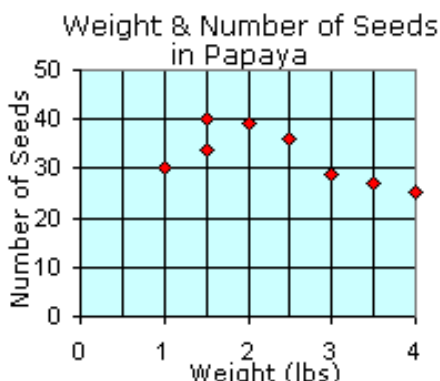
**Graph 1**



**Graph 2**

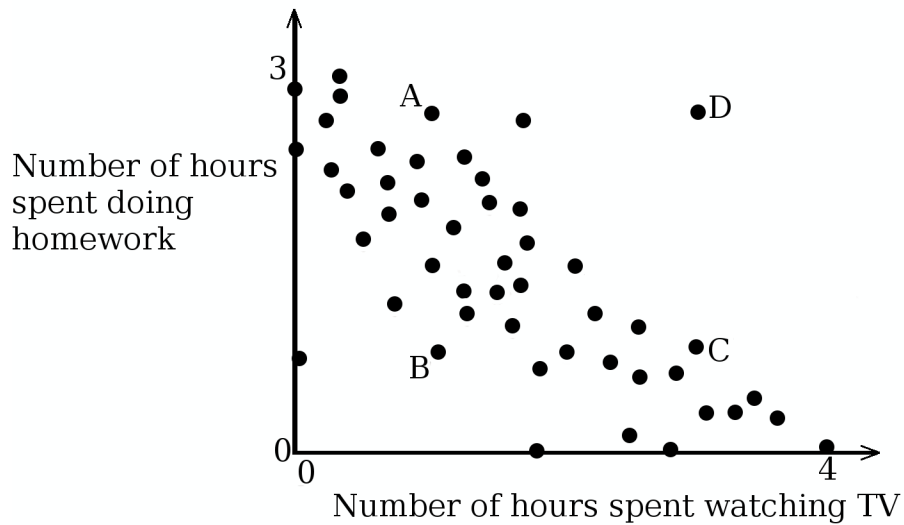


**Graph 3**



**Graph 4**

Roy was doing research for a research paper. He questioned students throughout his high school, asking them how much time they spent doing homework and how much time they spent watching TV the previous evening. The following scatter plot shows his results. Based on the information answer the questions that follow.



Choose the best of the 4 points, A, B, C, or D to represent the student's statements below.

12. "I worked on homework almost all night, I only had time to watch my favorite sitcom."
13. "Last night was about half and half for me"
14. "Last night didn't have anything on the screen I wanted to watch, and homework was so light, that I ended up going out."
15. Write a statement that correlates to the 4<sup>th</sup> point.

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.13.

## 4.14 Creating Pie Charts

### Objective

Here you will learn to create an accurate pie chart to display comparative data.

### Concept

Kalena's cheerleading squad is raising money for prom by selling candy at school football games. After a month of sales, the squad is running low on candy and decides to review the sales so far to help them decide what to order when they restock.

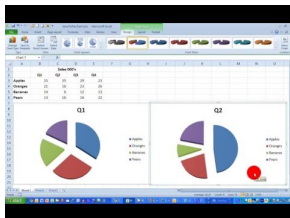
The table below describes the sales that the girls have recorded. How could the girls create a visual representation of the data so they can more easily present their findings to the purchasing committee? Ideally, they would like to order more of the item that is proving the most profitable, and so they want to present the data to the committee in a way that highlights this. We will review this question after the lesson.

**TABLE 4.87:**

Item	# of Sales
Popsicles \$1 ea	850
Chocolate Bars \$0.85 ea	1300
Bag of Lemon Drops \$1.25 ea	340
Ice Cream Bars \$1 ea	670

### Watch This

There are two videos applicable to this lesson, the first is a demonstration of creating a pie chart with the use of spreadsheet software, the second explains creation by hand.



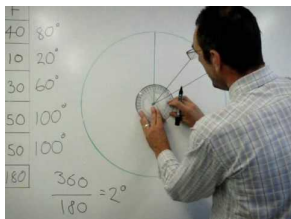
### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63840>

<http://youtu.be/IVIXbH4nczI> eoloughlin - How To... Draw a Simple Pie Chart in Excel 2010

Just for clarity, note that this video shows the same pie chart created *twice*, once incorrectly, and then once correctly, as stated in the first few seconds of the video.

**MEDIA**

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/63842>

[http://youtu.be/W60F\\_517PVY](http://youtu.be/W60F_517PVY) krmathsandscience - How to Construct a Pie Chart

**Guidance**

Pie charts are an excellent way to compare multiple values that make up parts of a whole. Each piece of the pie is called a **sector**, and each sector encompasses an angle that is proportional to the frequency of the data it represents. The formula relating the angle of a sector to frequency is:

$$\text{Sector Angle} = \frac{\text{Frequency of Data in the Sector}}{\text{Total Frequency of Data in Sample}} \times 360^\circ$$

In order to construct an accurate pie chart, you will need to calculate the sector angles for each of the categories or intervals in your sample, verifying that they total  $360^\circ$ .

Once you know the angles for each sector construct a circle and mark the sectors within it with lines running from the center of the circle to the edge (radii). Make sure that the angle between the lines is equal to the calculated sector angle for each category.

Finally you need to either label the sectors directly or create a key similar to the one in the concept question above so your audience can easily identify which sector corresponds to each category in your sample.

**Example A**

In Karen's school, there were 480 students in 1997, 540 students in 2000, 710 students in 2003, and 900 students in 2006. Construct a pie chart to represent the relative numbers of students each year.

Solution:

First, calculate the total number of students over all four categories (years):

$$\text{Total number of students} = 480 + 540 + 710 + 900 = 2,630$$

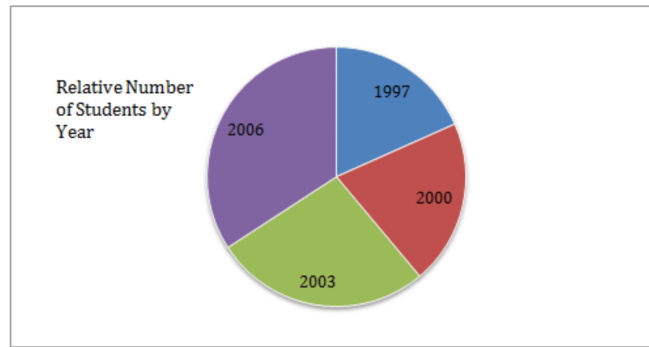
Now calculate the angle of each sector with  $\text{Sector Angle} = \frac{\text{Frequency of Data in the Sector}}{\text{Total Frequency of Data in Sample}} \times 360^\circ$

- 1997:  $\text{sector angle} = \frac{480}{2630} \times 360^\circ \rightarrow .183 \times 360^\circ \rightarrow \approx 66^\circ$
- 2000:  $\text{sector angle} = \frac{540}{2630} \times 360^\circ \rightarrow .205 \times 360^\circ \rightarrow \approx 74^\circ$
- 2003:  $\text{sector angle} = \frac{710}{2630} \times 360^\circ \rightarrow .270 \times 360^\circ \rightarrow \approx 97^\circ$
- 2006:  $\text{sector angle} = \frac{900}{2630} \times 360^\circ \rightarrow .342 \times 360^\circ \rightarrow \approx 123^\circ$

Verify that your degree measures total  $360^\circ$ :

$$66^\circ + 74^\circ + 97^\circ + 123^\circ = 360^\circ$$

Finally, construct your circle and draw the internal angles equal to the calculated sector angles, and color-code and/or directly label each sector:

**Example B**

Create a pie chart to display the data from the table below:

**TABLE 4.88:** Makes of Student Cars in Parking Lot

Ford	57
Chevrolet	49
Dodge	36
Toyota	27
Nissan	16
BMW	5
Mercedes	3
All Others	17

**Solution:** First total the number of cars in the entire population:

$$57 + 49 + 36 + 27 + 16 + 5 + 3 + 17 = 210$$

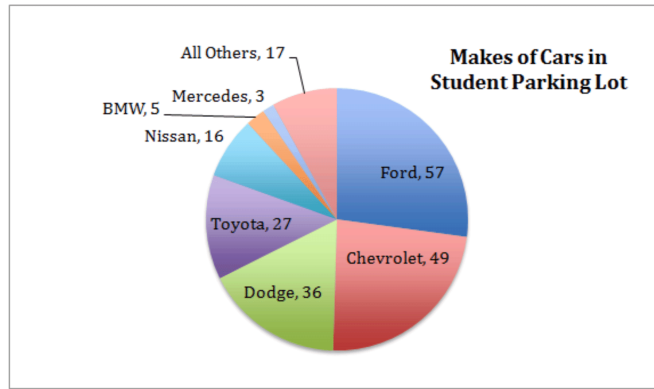
Now find the degree measure of each sector using:

$$\text{Sector Angle} = \frac{\text{Frequency of Data in the Sector}}{\text{Total Frequency of Data in Sample}} \times 360^\circ$$

- Ford:  $\text{Angle} = \frac{57}{210} \times 360^\circ \approx 98^\circ$
- Chevrolet:  $\text{Angle} = \frac{49}{210} \times 360^\circ = 84^\circ$
- Dodge:  $\text{Angle} = \frac{36}{210} \times 360^\circ \approx 62^\circ$
- Toyota:  $\text{Angle} = \frac{27}{210} \times 360^\circ \approx 46^\circ$
- Nissan:  $\text{Angle} = \frac{16}{210} \times 360^\circ \approx 27^\circ$
- BMW:  $\text{Angle} = \frac{5}{210} \times 360^\circ \approx 9^\circ$
- Mercedes:  $\text{Angle} = \frac{3}{210} \times 360^\circ \approx 5^\circ$
- All Others:  $\text{Angle} = \frac{17}{210} \times 360^\circ \approx 29^\circ$

Verify that the total is  $360^\circ$ :  $98^\circ + 84^\circ + 62^\circ + 46^\circ + 27^\circ + 9^\circ + 5^\circ + 29^\circ = 360^\circ$

Construct a circle with sectors representing each degree measure and label directly or create a key:



**Example C**

Use a spreadsheet or compass and protractor to create two related pie charts of the data in the table below. If using a modern spreadsheet, create a 3-D graph. Highlight the data from the soccer participants in both graphs.

**TABLE 4.89:**

Sport	Count of Middle School Participants	Count of High School Participants
Football	186	279
Volleyball	28	57
Soccer	66	92
Track	82	124

**Solution:** By now you are familiar with creating charts using a pencil and paper, so let’s walk through creating a chart in a modern spreadsheet.

If you do not have spreadsheet software on your computer, you can download the *free* ‘Open Office Calc’ spreadsheet software from <http://www.openoffice.org/> . The process described here is essentially the same in any modern spreadsheet software such as Open Office Calc, Microsoft Excel or Numbers.

First, highlight and copy the two columns of data under ‘Sport’ and ‘Count of Middle School Participants’ from the table in the question, include the column headers.

Now open a blank spreadsheet in your software and paste the two columns of data.

Highlight the data in the spreadsheet (including the headers), and either choose ‘Charts : Pie’ from the toolbar or click the icon that looks like a bar graph and choose ‘Pie.’

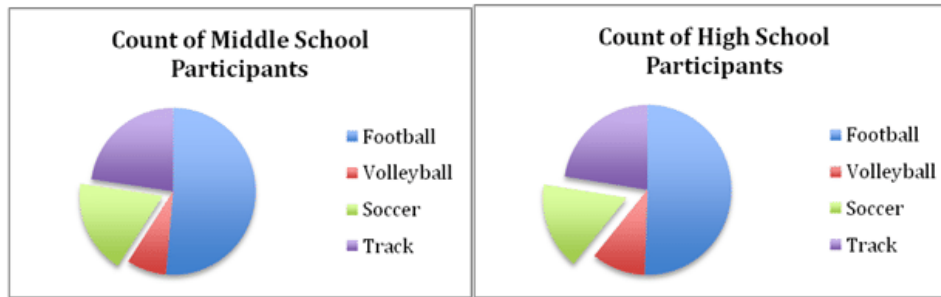
The data will be immediately converted to a pie chart for you! Now pick a 3-D style (since that is what was specified in the question), and press ‘enter’ to get a chart like the one below.

To highlight the soccer data, simply select the sector representing the soccer participants and drag it away from the center of the circle just slightly. Finally, select the entire chart in the spreadsheet and copy/paste it into your answer document or print it out to turn it in.

To create the chart for the high school data, just copy the data from the single ‘Count of High School Participants’ category and paste it right overtop of the Middle School column in your spreadsheet and repeat the steps above to convert the revised table into a chart.

The two final products should look something like these:





### Concept Problem Revisited

Create a pie chart that compares the income from each product for the cheerleading squad. Use the data in the table of candy sales.

In the example problems, the first step was to total the number of items sold. However, in this problem, we need to compare the *dollar value* of the items rather than just the *number* of items. That means we need to first evaluate the dollar value of each product sale, and then calculate the angle of each slice based on a comparison of the dollar value of each product with the total income from sales. Just to keep things neat, let's add another column to the original table called "dollar value", and another row at the bottom for the total.

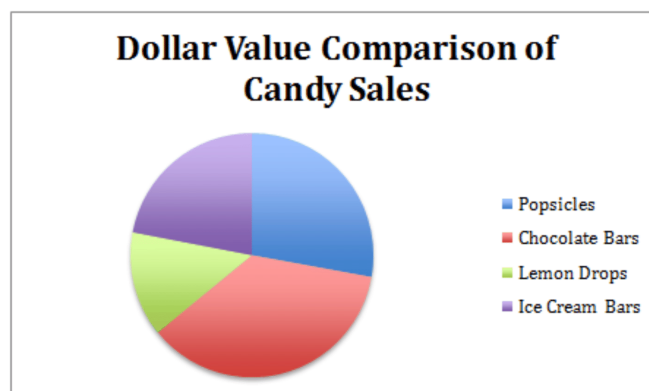
**TABLE 4.90:**

Item	# of Sales	\$ Value
Popsicles \$1 ea	850	\$850.00
Chocolate Bars \$0.85 ea	1300	\$1,105.00
Bag of Lemon Drops \$1.25 ea	340	\$425.00
Ice Cream Bars \$1 ea	670	\$670.00
TOTAL	3160	\$3,050.00

Now we can use the formula to calculate the angle of each slice:

- Popsicles:  $Angle = \frac{\$850}{\$3,050} \times 360^\circ \approx 100^\circ$
- Chocolate Bars:  $Angle = \frac{\$1,105}{\$3,050} \times 360^\circ \approx 130^\circ$
- Lemon Drops:  $Angle = \frac{\$425}{\$3,050} \times 360^\circ \approx 50^\circ$
- Ice Cream Bars:  $Angle = \frac{\$670}{\$3,050} \times 360^\circ \approx 79^\circ$

Finally, we construct our circle and mark the divisions of the sectors based on the angles we have calculated, label the sectors, and label the graph.



## Vocabulary

A *sector* is a single 'pie slice' in a circle graph.

The *whole relationship* is represented by the entire circle.

## Guided Practice

1. Larry Bird was a well know basketball player. He played for the Boston Celtics. Use the following information to create a pie chart.

**TABLE 4.91:**

Season:	# of Points Scored
1979-1980	1745
1980-1981	1741
1981-1982	1761
1982-1983	1867
1983-1984	1908
1984-1985	2295
1985-1986	2115
1986-1987	2076
1987-1988	2275
1988-1989	116
1989-1990	1820
1990-1991	1164
1991-1992	908

2. What percent of his career points came from the 1988 season?

3. What percent of his career points came from the 1980 season?

The following table shows the grades achieved by 30 pupils in their end of year exam.

**TABLE 4.92:**

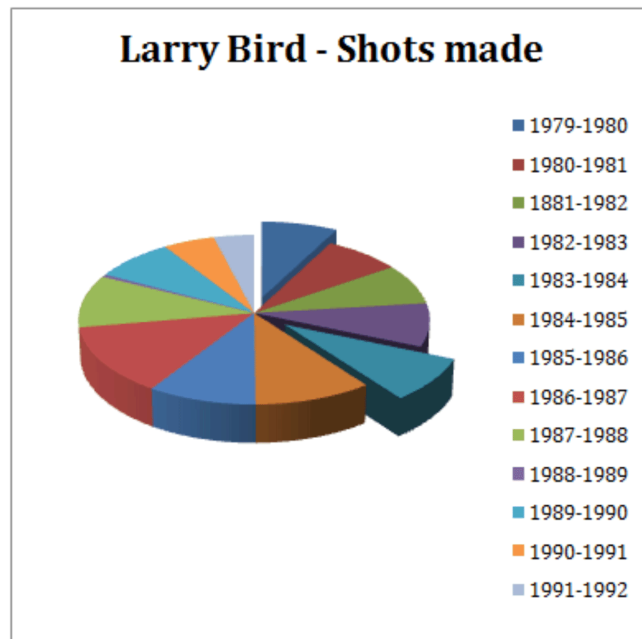
Grade	A	B	C	D	F
Frequency	8	10	7	3	3

4. Based on the number of pupils, how many degrees of a circle graph are allotted per student? What percentage of the pie cart is allotted to "C" grades?

5. Create a pie chart for the information.

## Solutions:

1.



2. First we must find the total number of baskets Larry made in the portion of his career represented by the chart. The answer is 21,792. This number represents 100 percent of our chart. To find what percentage of those shots were made in 1988, we write an algebraic word problem that looks this:

What percent of 21,792 is 2275? As an equation:  $\boxed{?} \times 21,792 = 2275 \rightarrow \boxed{?} = \frac{2275}{21792} \rightarrow \boxed{?} = 10.4\%$

Percent of career shots made in 1988 is 10.4%

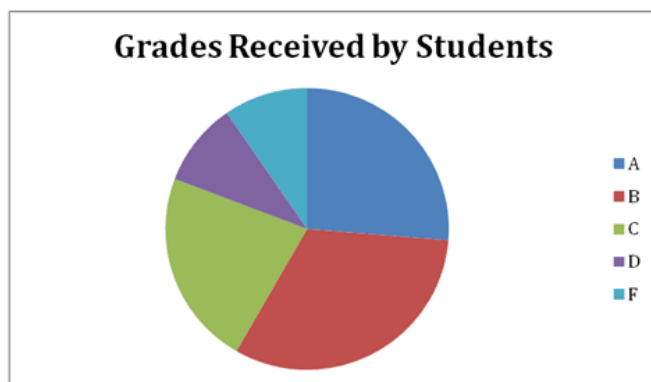
3. Calculate this the same way:  $x\% \times 21792 = 1741 \rightarrow \frac{1741}{21792} = .08 \rightarrow x = 8\%$

Percent of shots made in his career in 1980 is 8%

4. There are 31 total grades, 7 of which are “C’s”. To find percentage, just divide:  $\frac{7}{31} = .225$ , and then multiply by 100 to get percentage. **“C” grades make up 22.5% of the chart.**

5. Your pie chart should look something like the image below, with sectors measuring:

- Grade A:  $\frac{8}{31} = .26 = 26\%$  Convert to degrees:  $.26 \times 360^\circ = 94^\circ$
- Grade B:  $\frac{10}{31} = .32 = 32\%$  Convert to degrees:  $.32 \times 360^\circ = 115^\circ$
- Grade C:  $\frac{7}{31} = .225 = 23\%$  Convert to degrees:  $.225 \times 360^\circ = 81^\circ$
- Grade D:  $\frac{3}{31} = .097 = 10\%$  Convert to degrees:  $.097 \times 360^\circ = 35^\circ$
- Grade F:  $\frac{3}{31} = .097 = 10\%$  Convert to degrees:  $.097 \times 360^\circ = 35^\circ$



**Practice**

1. What is each part of a pie chart called?
2. What part of the circle represents the whole relationship?
3. Based on the data in the table below, which candy sold the most, and which the least?

**TABLE 4.93:**

Candy Bar	Number sold at the school store	Percentage of Total Sales	Number of Degrees on a Pie Chart
Heath	151		
M	M	191	
Snickers	61		
Skittles	107		
Almond Joy	91		

4. Fill in the chart above with percentages of each candy type sold.
5. Fill in the chart with degrees of a circle required to represent this amount on a pie chart.
6. Create a Pie Chart to represent this data.
7. Based on the data in the table below, which stock has the greatest potential for making money for the investor who owns the stock?

**TABLE 4.94:**

Company	Shares Owned	Percentage of Total Portfolio	Number of Degrees on a Pie Chart
Hostess	8		
Pepsi	11		
Dell	5		
Conoco	7		
Ford Motor	19		

8. Fill in the chart with relative percentages of each type of stock.
9. Fill in the chart with degrees of a circle required to represent this amount on a pie chart.
10. Create a Pie Chart to represent this data.

Use the data presented in questions 11-15 to create pie charts for each:

11. Julie runs for one hour per day, reads for two and sleeps nine. She spends about two hours eating, at least one hour on the phone with friends. She hangs out with her family on average 4 hours a day, and spends at least five hours a day studying.
12. Mrs. Garcia makes \$1200.00 a month. She puts 10% in savings, spends 20% on her car payment and insurance, and another 20% on groceries. She likes clothes, so 10% of her income goes towards her wardrobe. Of her remaining money she spends 30% on her mortgage, and the remainder on miscellaneous expenses. Determine exactly how much Mrs. Garcia spends in each category. What percent of her income goes to miscellaneous expenses?
13. Katie earned 500.00 doing odd jobs for people in her neighborhood. She spent  $\frac{1}{16}$  of her money on the movies, she spent  $\frac{3}{8}$  of her money going out with friends. She spent  $\frac{1}{2}$  on clothes, and the remainder on books. How much money did Katie spend on movies? What percentage was spent on books?

14. The state of Colorado receives 28 inches of precipitation a year. The winter is when it gets most of it, but it does not see it until it melts and runs off in the spring. However, inches are counted when they accumulate, and so they represent precipitation for Colorado based on a 4 season year. Colorado receives  $\frac{3}{5}$  of its precipitation in the winter,  $\frac{3}{10}$  in the spring and the rest in the summer and fall months.

15. Students were preparing to go on a field trip, and their teacher let them choose the destination. There were 36 students in the class. 44.4% choose the Nature Preserve, 25% chose an Art Gallery. Half as many students wanted to go to the Symphony as the Nature Preserve, and the rest wanted to go to the Museum of Nature and Science. How many more students were there that wanted to go to the Nature Preserve than wanted to go to the Museum of Nature and Science?

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.14.

## 4.15 Interpreting Pie Charts

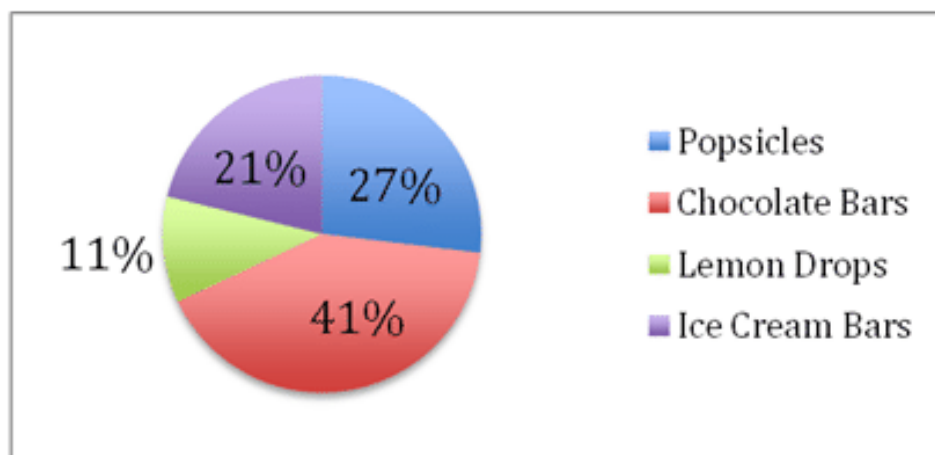
### Objective

Here you will learn to read and evaluate the comparative data on a pie chart.

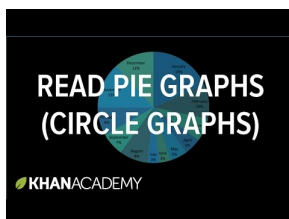
### Concept

Kalena's cheerleading squad is raising money for prom by selling candy at school football games. After a month of sales, the squad is running low on candy and decides to review the sales so far to help them decide what to order when they restock.

The pie chart below describes the number of sales that the girls have recorded for each item. If you know that they sold 850 fruit popsicles, how could you calculate the number of chocolate bars or bags of lemon drops they sold? We will review this question after the lesson.



### Watch This



#### MEDIA

Click image to the left or use the URL below.

URL: <https://www.ck12.org/flx/render/embeddedobject/62548>

<http://youtu.be/4JqH55rLGKY> KhanAcademy - Reading Pie Chart Graphs

### Guidance

Reading a pie chart is relatively simple, which is one of the primary values of a pie chart. The only real trick to it is to become very familiar with the *sector angle formula* and to practice using it to deduce the value(s) of unspecified

data.

Recall that each piece of the pie is called a **sector**, and each sector encompasses an angle that is proportional to the frequency of the data it represents. The formula relating the angle of a sector to frequency is:

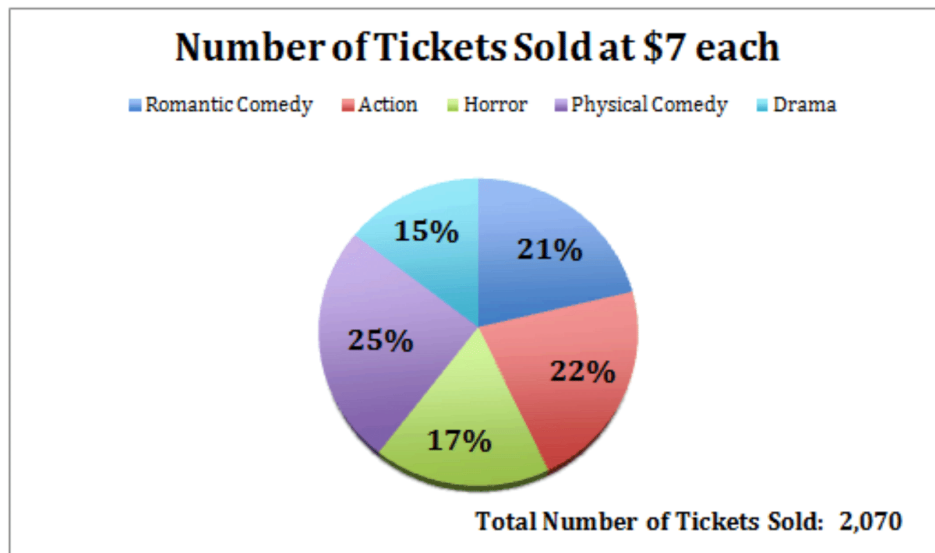
$$\text{Sector Angle} = \frac{\text{Frequency of Data in the Sector}}{\text{Total Frequency of Data in Sample}} \times 360^\circ$$

Because the primary skill here is just familiarity that comes with practice, let's jump right into the examples.

### Example A

Given the pie chart below, identify or calculate the following:

- The total number of tickets to comedy-themed movies.
- The dollar value of tickets sold to horror movies.
- The percentage of tickets sold to action and romantic comedies together.



Solution:

a. To calculate the total number of comedy movie tickets, we need to multiply the decimal equivalent of each of the two comedy categories by the total number of tickets sold to learn how many of each were sold, then find the sum.

i.  $.21 \times 2070 = 435$  *romantic comedy tickets*

j.  $.25 \times 2070 = 518$  *physical comedy tickets*

k.  $435$  *romantic comedy* +  $518$  *physical comedy* =  $953$  *Total Comedy Tickets*

b. To find the dollar value of the horror movies, first multiply the decimal equivalent of the horror movie percentage, and multiply the result by \$7, the price of each ticket.

l.  $.17 \times 2070 = 352$  *horror movie tickets*

m.  $352 \times \$7 = \$2,464$  *in horror movie ticket sales*

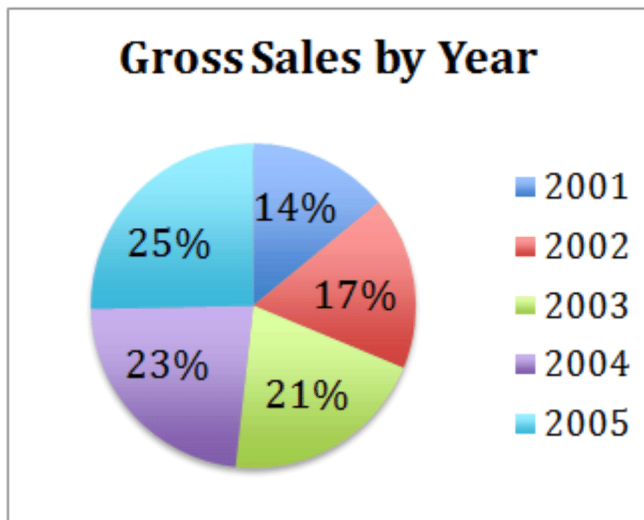
c. To find the percentage of action and romantic comedy tickets together, simply sum the given percentages.

n.  $22\% + 21\% = 33\%$  *total action and romantic comedy tickets*

### Example B

Given the pie chart below, calculate the actual gross sales in years 2001, 2002, 2003, and 2004.

Gross sales in 2005 were \$94,000



**Solution:** First we need to calculate the total gross sales for all five years. We know that the \$94,000 in gross sales from 2005 represents 25% of the total, so we can calculate:

$$.25 \times x = \$94,000 \rightarrow \frac{\$94,000}{.25} = \$376,000$$

Now we can simply multiply the total gross sales by the percentage represented by each year to get the estimated dollar value for each category:

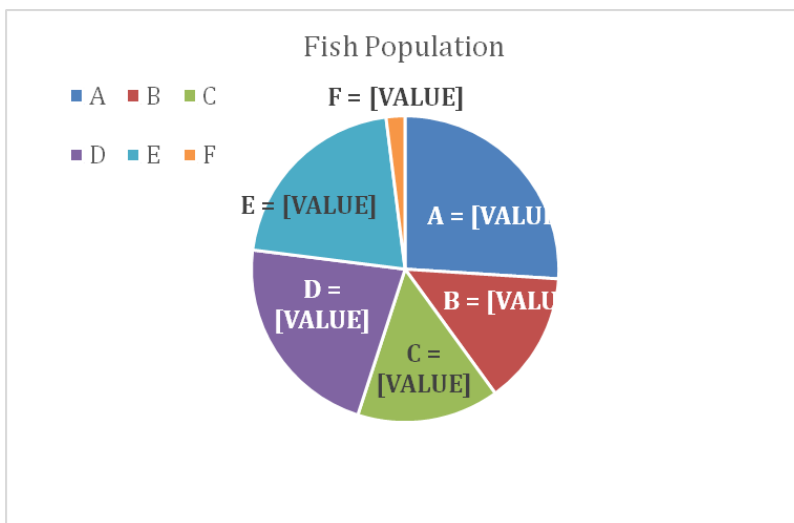
- 2001: 14% of \$376,000 = \$52,640
- 2002: 17% of \$376,000 = \$63,920
- 2003: 21% of \$376,000 = \$78,960
- 2004: 23% of \$376,000 = \$86,480

Verify that the total values add up:  $\$52,640 + \$63,920 + \$78,960 + \$86,480 + \$94,000 = \$376,000$

**Example C**

Evaluate the pie chart below and answer the questions:

- If the population of Lake C is 27,000 fish, what is the total fish population of all lakes on the chart?
- What is the ratio of the fish population in lakes A and C?
- What is the combined population of lakes D and E?





**Solution:**

a. The population of lake C is 27,000 fish, and represents 15% of the whole, use the data to set up an equation:

$$.15x = 27,000 \rightarrow x = \frac{27000}{.15} \rightarrow x = 180,000$$

**The combined population of all lakes on the chart is 180,000 fish.**

b. Lake A represents 26% of the total, and Lake C 15%.

$$\text{The ratio is } \frac{26}{15} = 1.73 : 1$$

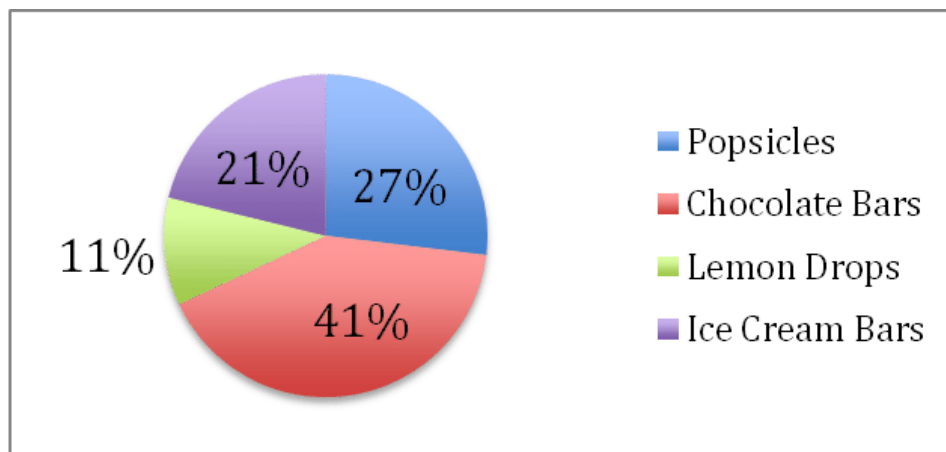
c. Lake D is 22% of 180,000 and Lake E is 21% of 180,000.

Together they represent  $22\% + 21\% = 43\%$  of the total.

$$43\% \times 180,000 = 77,400 \text{ fish in lakes D and E together.}$$

**Concept Problem Revisited**

The pie chart below describes the sales that the girls have recorded as numbers of sales. If you know that they sold 850 fruit popsicles, how could you calculate the number of chocolate bars or bags of lemon drops they sold?



This one should be a piece of... pie... now! 850 fruit popsicles represents 27% of the total. Set up an equation:

$$27\%x = 850 \rightarrow x = \frac{850}{.27} \rightarrow 3,148 \text{ total sales}$$

Chocolate Bars represent 41% of the 3,148 total sales:

$$.41 \times 3148 = 1291 \text{ chocolate bars}$$

Lemon Drops represent 11% of the 3,148 sales:

$$.11 \times 3148 = 346 \text{ lemon drop bags}$$

**Vocabulary**

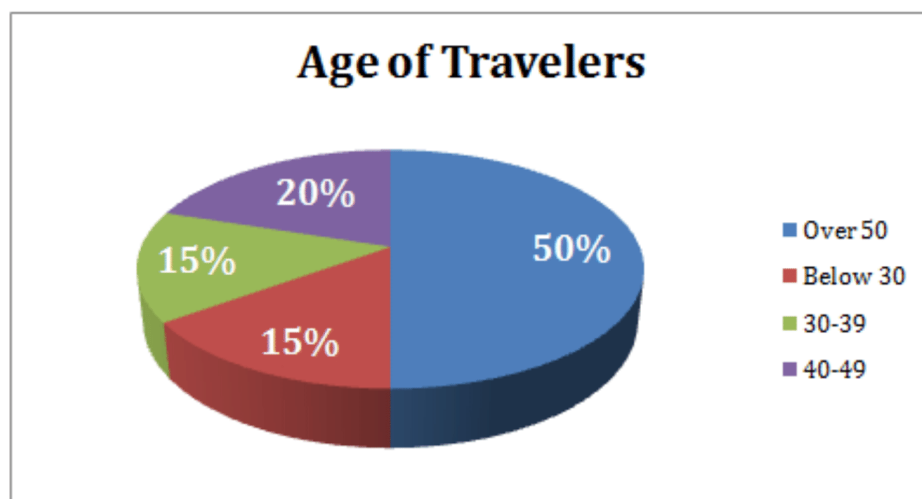
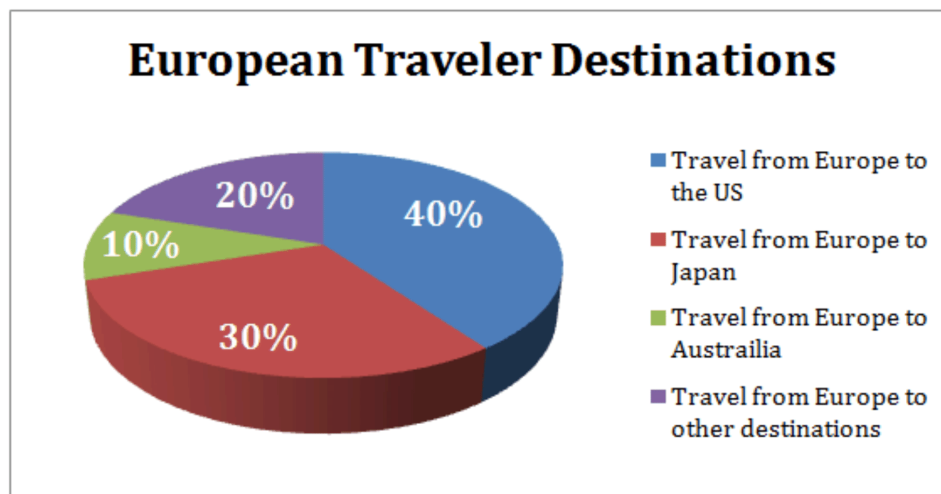
A **sector** is a single portion of a pie or circle graph (one “slice”)

The **sector angle formula** is used to calculate how many degrees of the circle should be allocated to a given value.

The formula is:  $\text{Sector Angle} = \frac{\text{Frequency of Data in the Sector}}{\text{Total Frequency of Data in Sample}} \times 360^\circ$

### Guided Practice

The two charts below show the distribution of overseas tourist traffic from Europe to other countries. They show the distribution by country and by age, respectively.



1. What percentage of tourists went to either the US or Australia?
2. What is the ratio of European tourists that went to the USA to the number of European tourists who were below the age of 30?
3. If, among the “other destinations”, India accounted for 25% of the travel, and it is known that 25 Europeans went to India, how many 30-39 year old Europeans went abroad in that year?
4. Based on the data, about how many 40-49 year olds traveled to Japan?

### Solutions:

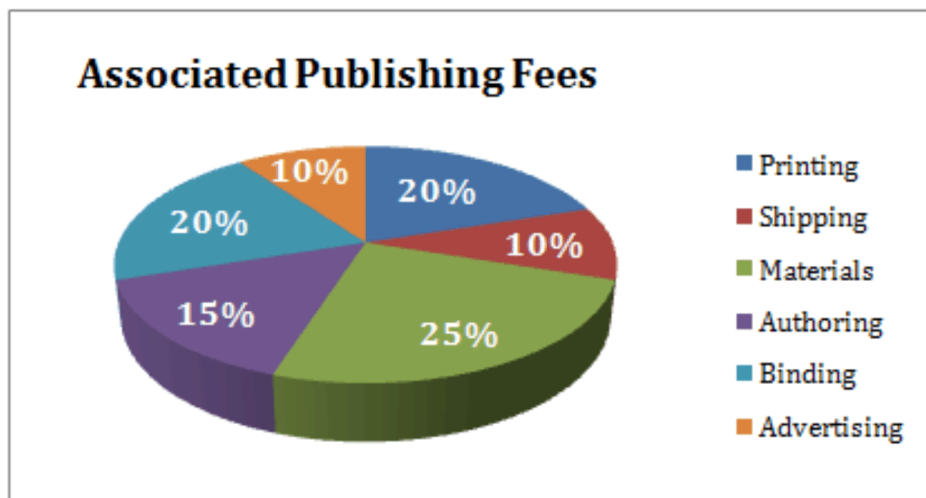
1. We can see from the top chart that U.S. travelers (in blue) accounted for 40% of the travel, and Australian travelers (green) represented 10%. The total percentage who went to either country is  $40\% + 10\% = 50\%$  of travelers.
2. 40% of travelers went to the U.S, and 15% were under 30 (bottom graph, in red). **The ratio is 40:15 or 8:3.**
3. If India accounts for 25% of the 20% of travelers who went to “other destinations”, then it accounts for 5% of the total travelers ( $25\% \times 20\% = 5\%$ ). If 25 Europeans went to India, then we know 25 travelers represents

5% of the total. We can calculate the total number of European travelers to be:  $5\% \times x = 25 \rightarrow x = \frac{25}{.05} \rightarrow x = 500$  total travelers. We then look at the *Age of Travelers* chart and see that 30-39 year olds accounted for 15% of the travel that year. Multiply 15% by 500 to find our answer, which is **75 travelers**.

4. We know there are a total of 500 travelers. We can look at the number of total travelers who went to Japan, which is 30% of the 500, or **150 people who went to Japan**. Out of that 150, we can guess that 20% (the percentage of 40-49 year old travelers, from the bottom chart) of them were between the age of 40-49.  $20\% \times 150 = 30$  **people**.

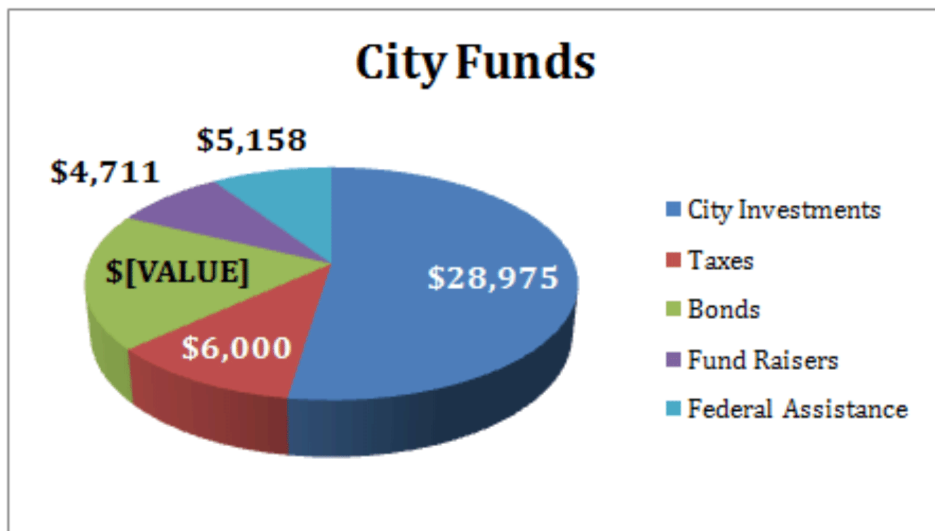
## Practice

The following pie-chart shows the percentage distribution of the expenses incurred in publishing a statistics math book. Study the Pie Chart and answer the questions below.



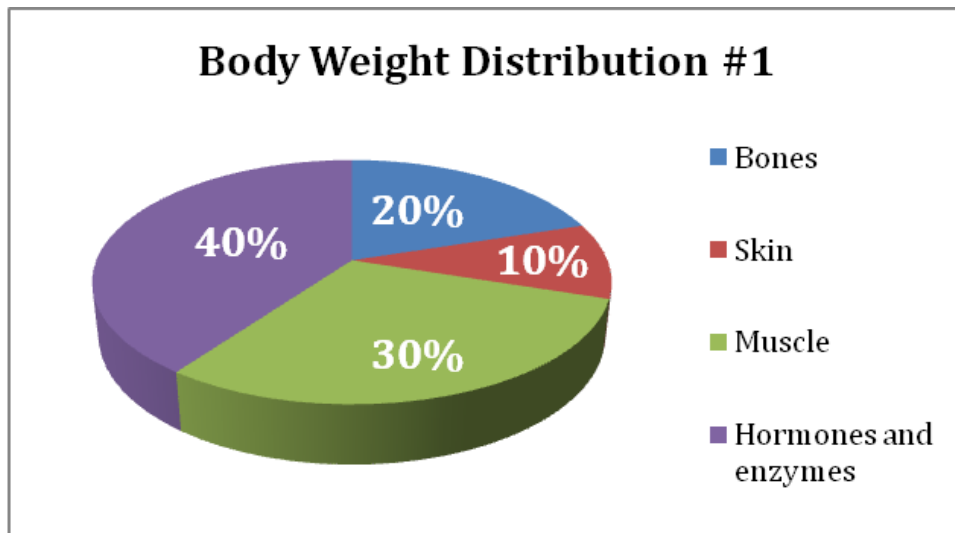
- If the publisher paid \$30,600 for printing, then how much did the publisher pay the author to write the book?
- What is the measurement of the angle of the sector that corresponds to the authoring expense?
- If the book is sold at 20% above the actual cost per book, and the marked price is \$180.00, what is the cost of the paper used in a single copy of the book?
- If 5500 copies of the book are published, and shipping on them amounts to \$82,500, what should the selling price of the book be so that the publisher can earn a profit of 25%?
- Authorship of the book is less than the printing cost by what percentage?
- If the difference between two of the expenditures in the chart are represented by 54 degrees, then which combination(s) of two expenditures could they be?
- If the cost of paper is \$56,250 for printing an edition of the book, then what was the advertising cost?
- Identify two expenditures that together have a central angle of 108 degrees.

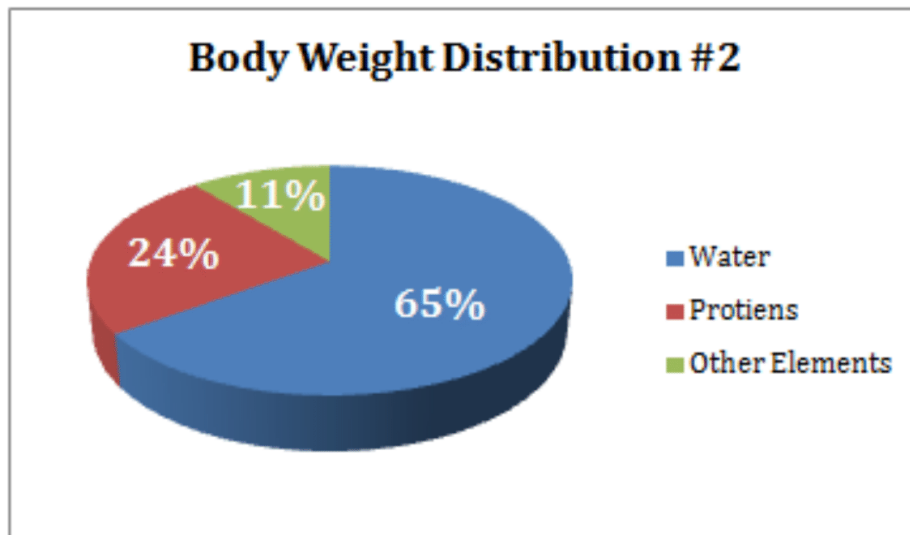
The following pie chart shows the sources of funds that will be collected by Downtown Development to beautify the city. Use the chart to answer the questions that follow.



9. Nearly 20% of the funding will come from which source? 55290
10. If the city could only pull \$9,725 from bonds, by what percent should it increase the amount it borrows against its investments to make up the difference?
11. If the bonds are managed by a third party that collects a 10% commission, how much would the bonds have actually been worth, before being added to the budget represented in the chart above?
12. What is the degree measure corresponding to the Taxes collected?
13. What is the approximate ratio of the funds to be collected through Bonds and through City Investments?

The following charts give information about how weight is distributed throughout the human body, according to different components. Answer the questions below based on the pie charts.





14. What percentage of proteins is equivalent to the weight of skin?
15. How much of the human body is neither made of bones or skin?
16. What is the ratio of the weight of proteins in the muscles to that of the weight of proteins in the bones?

### Answers for Explore More Problems

To view the Explore More answers, open this [PDF file](#) and look for section 4.15.

Students were given a quick review of data collection, and were taught about relative and cumulative frequencies before the bulk of the chapter detailed the process of constructing and interpreting various common data visualizations. Histograms, box plots, stem-and-leaf diagrams, scatter plots, line graphs, pie (circle) charts and frequency polygons were each separately taught.

After completing the lessons in this chapter, students should feel confident in their ability to create charts of specific types to display data and to read data from charts to draw conclusions.

---

## 4.16 References

1. reynermmedia. <https://www.flickr.com/photos/89228431@N06/11322953266> .
2. ~Pawsitive~Candie\_N. <https://www.flickr.com/photos/scjn/4269256333> .
3. Paul Reynolds. <https://www.flickr.com/photos/bigtallguy/139143816> .
4. flash.pro. <https://www.flickr.com/photos/flashpro/4156535452> .
5. Daniel Oines. <https://www.flickr.com/photos/dno1967b/5543841808> .
6. . . CC BY-NC-SA
7. CK-12 Foundation. . CCSA
8. CK-12 Foundation. . CCSA
9. Marc Berry Reid. <https://www.flickr.com/photos/marcberryphotos/2983082444> .
10. CK-12 Foundation. . CCSA
11. CK-12 Foundation. . CCSA
12. CK-12 Foundation. . CCSA
13. CK-12 Foundation. . CCSA
14. . . CC BY-NC-SA
15. . . CC BY-NC-SA
16. . . CC BY-NC-SA
17. . . CC BY-NC-SA
18. . . CC BY-NC-SA
19. CK-12 Foundation. . CCSA
20. CK-12 Foundation. . CCSA
21. CK-12 Foundation. . CCSA
22. CK-12 Foundation. . CCSA
23. CK-12 Foundation. . CCSA
24. CK-12 Foundation. . CCSA
25. CK-12 Foundation. . CCSA
26. CK-12 Foundation. . CCSA
27. CK-12 Foundation. . CCSA
28. CK-12 Foundation. . CCSA
29. CK-12 Foundation. . CCSA
30. . . CC BY-NC-SA
31. CK-12 Foundation. . CCSA
32. CK-12 Foundation. . CCSA
33. CK-12 Foundation. . CCSA
34. CK-12 Foundation. . CCSA
35. CK-12 Foundation. . CCSA
36. CK-12 Foundation. . CCSA
37. CK-12 Foundation. . CCSA
38. CK-12 Foundation. . CCSA
39. CK-12 Foundation. . CCSA
40. . . CC BY-NC-SA
41. CK-12 Foundation. . CCSA
42. . . CC BY-NC-SA
43. CK-12 Foundation. . CCSA
44. CK-12 Foundation. . CCSA
45. CK-12 Foundation. . CCSA

46. CK-12 Foundation. . CCSA
47. CK-12 Foundation. . CCSA
48. CK-12 Foundation. . CCSA
49. CK-12 Foundation. . CCSA
50. CK-12 Foundation. . CCSA
51. CK-12 Foundation. . CCSA
52. CK-12 Foundation. . CCSA
53. CK-12 Foundation. . CCSA
54. CK-12 Foundation. . CCSA
55. CK-12 Foundation. . CCSA
56. CK-12 Foundation. . CCSA
57. CK-12 Foundation. . CCSA
58. CK-12 Foundation. . CCSA
59. CK-12 Foundation. . CCSA
60. CK-12 Foundation. . CCSA
61. CK-12 Foundation. . CCSA
62. CK-12 Foundation. . CCSA
63. CK-12 Foundation. . CCSA
64. CK-12 Foundation. . CCSA
65. CK-12 Foundation. . CCSA
66. CK-12 Foundation. . CCSA
67. CK-12 Foundation. . CCSA
68. CK-12 Foundation. . CCSA
69. CK-12 Foundation. . CCSA
70. CK-12 Foundation. . CCSA
71. CK-12 Foundation. . CCSA
72. CK-12 Foundation. . CCSA
73. CK-12 Foundation. . CCSA
74. CK-12 Foundation. . CCSA
75. CK-12 Foundation. . CCSA
76. CK-12 Foundation. . CCSA
77. . . CC-BY-NC-SA
78. . . CC BY-NC-SA
79. . . CC BY-NC-SA
80. . . CC-BY-NC-SA
81. . . CC-BY-NC-SA
82. . . CC-BY-NC-SA
83. . . CC BY-NC-SA
84. . . CC-BY-NC-SA
85. . . CC-BY-NC-SA
86. . . CC-BY-NC-SA
87. . . CC-BY-NC-SA
88. . . CC-BY-NC-SA
89. . . CC BY-NC-SA
90. . . CC-BY-NC-SA
91. . . CC BY-NC-SA
92. . . CC BY-NC-SA
93. . . CC-BY-NC-SA
94. . . CC-BY-NC-SA
95. . . CC-BY-NC-SA
96. . . CC-BY-NC-SA

97. . . CC-BY-NC-SA
98. . . CC-BY-NC-SA
99. CK-12 Foundation. . CCSA
100. CK-12 Foundation. . CCSA
101. . . CC BY-NC-SA
102. CK-12 Foundation. . CCSA
103. CK-12 Foundation. . CCSA
104. . . CC BY-NC-SA
105. CK-12 Foundation. . CCSA
106. CK-12 Foundation. . CCSA
107. CK-12 Foundation. . CCSA
108. . . CC BY-NC-SA
109. CK-12 Foundation. . CCSA
110. CK-12 Foundation. . CCSA
111. CK-12 Foundation. . CCSA
112. CK-12 Foundation. . CCSA
113. CK-12 Foundation. . CCSA
114. CK-12 Foundation. . CCSA
115. . . CC BY-NC-SA
116. CK-12 Foundation. . CCSA