EVALUATING DATA

In this unit, you will be evaluating different types of data. You will analyze data in tables, compare the same data in different types of graphs, investigate scatter plots and their lines of best fit, and interpret histograms. You will determine if data is symmetrical or skewed. You will evaluate types of biased and unbiased samples of data that is representative of a larger group and examine how data and graphs can be presented in misleading ways.

Same Data in Different Graphs

Scatter Plots

Histograms

Sampling

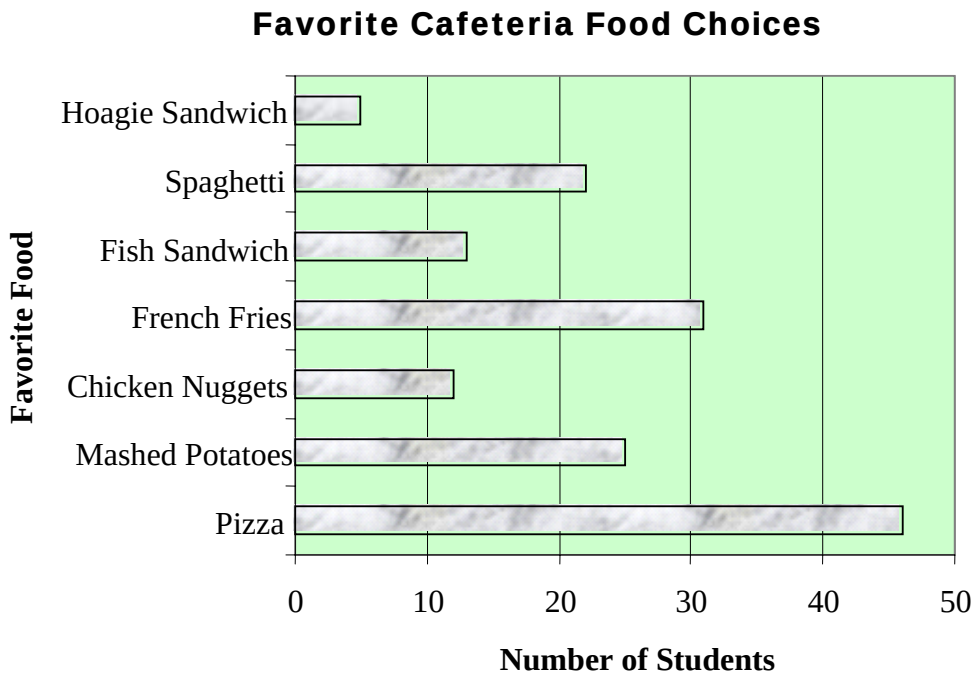
Misleading Statistics

Same Data in Different Graphs

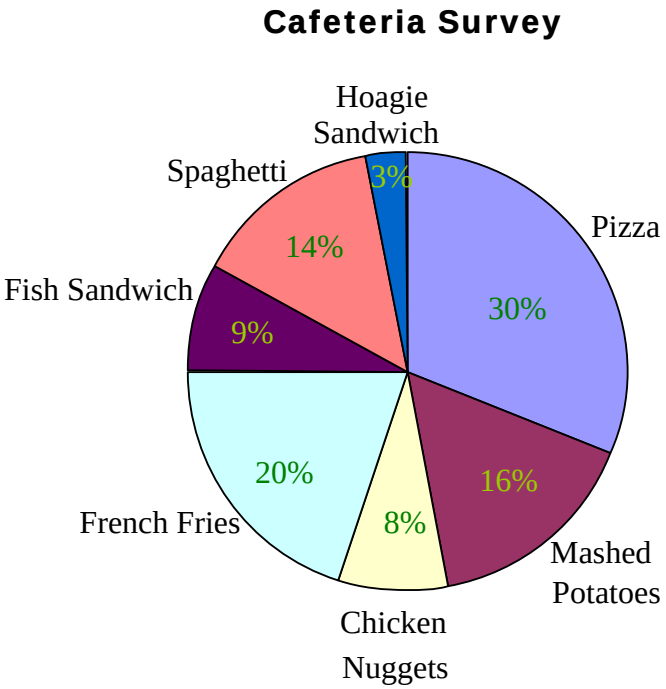
Jennifer surveyed the ninth grade students to determine the favorite food in the school cafeteria with these results: Pizza (46), Mashed Potatoes (25), Chicken Nuggets (12), French Fries (31), Fish Sandwich (13), Spaghetti (22), Hoagie Sandwich (5). She made a table to display her results.

Jennifer then decided to compare her results in two different types of graphs, a bar graph and a circle graph.

Which graph would you use to display the results to the student body?



In the **bar graph**, the **actual number** of students surveyed can be determined be reading the number represented by each bar.



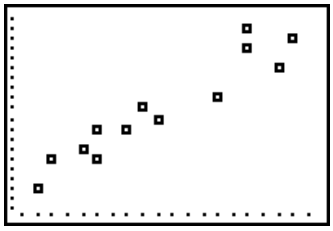
In the **circle graph**, the **percent of students** is shown rather than the number of students.

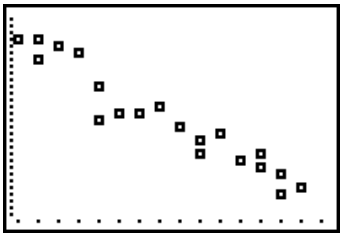
In both graphs, pizza is clearly the favorite choice of the students.

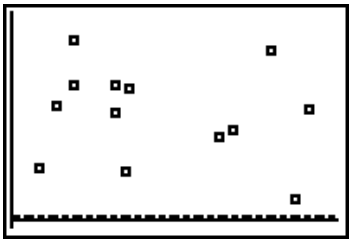
Scatter Plots

Scatter plots are an easy way of determining if there is a relationship between two variables. This relationship is called a **correlation**. A correlation is based on the slope of the line of best fit, a line that is drawn through the data and represents the overall trend of the data.

There are three possible types of correlation: a) positive, b) negative, or c) no correlation. The illustrations below show the graph of each correlation.

a) 

b) 

c) 

In graph "a", notice how the points cluster in a rise to the right. Recall from a previous lesson that this suggests a **positive slope**. Graph "b" shows points that cluster in a fall to the right, which suggests a **negative slope.** Graph "c" shows no cluster pattern and suggests the two variables have **no relationship** to each other.

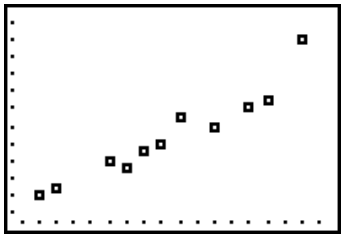
Let's take a look at a few examples and determine if each situation has a positive correlation, a negative correlation, or no correlation.

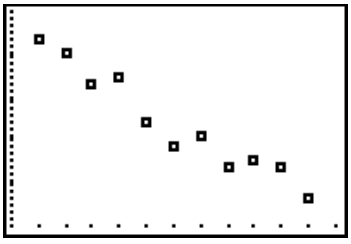
Example 1: Determine which scatter plot represents each situation.

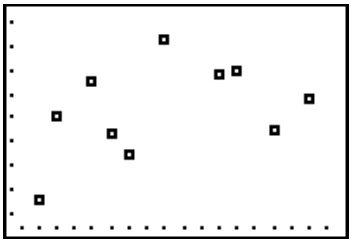
a) your height and your hourly wage

b) your height and your shoe size

c) your age and the time needed to run 100 yards

1) 

2) 

3) 

Scatter plot 1 shows a strong positive correlation. A **positive correlation** occurs when **both variables increase**. As you grow taller, your shoe size increases; therefore plot 1 represents situation "b".

Scatter plot 2 shows a strong negative correlation. A **negative correlation** occurs when **one variable increases as the other variable decreases**. In situation "c", as your age increases, the time it takes you to run 100 yards decreases. (Consider the time period from birth through young adulthood.)

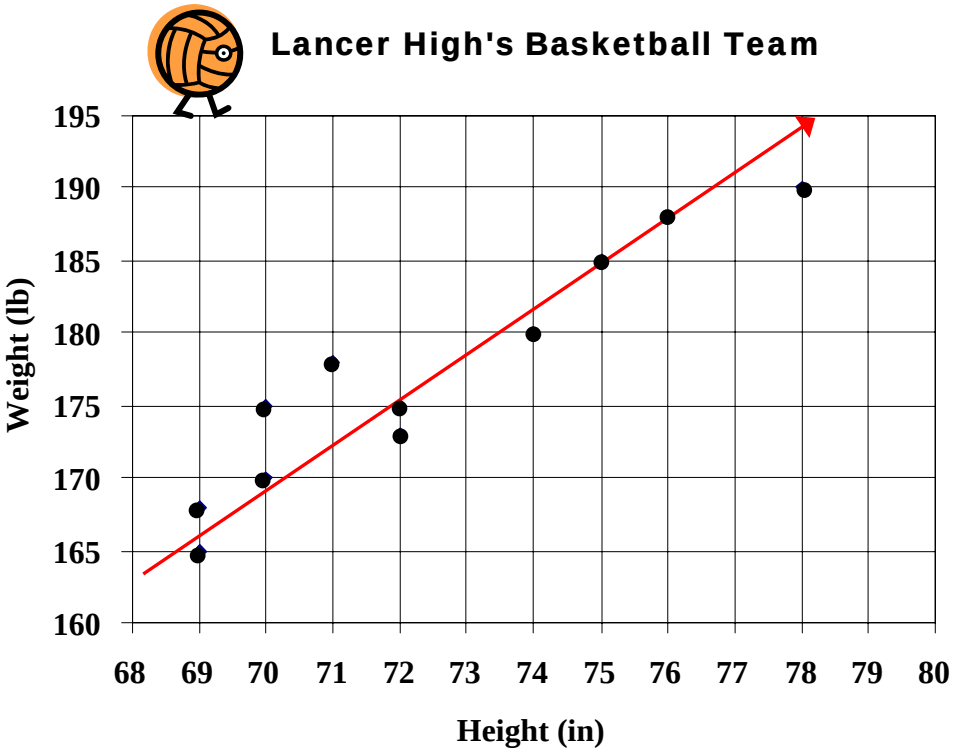
The third scatter plot shows **no correlation** because the data points are **randomly scattered**. Your height has no relationship with your hourly wage; therefore, this plot represents situation a.

Example 2: Use the data shown below to make a scatter plot of the weight and height of each member of Lancer High's basketball team.

Height (in)	Weight (lb)
70	170
69	165
72	175
74	180
75	185
70	175
69	168
72	173
71	178
78	190
76	188
69	165

Step 1: Make a scatter plot of the data pairs. The points on the scatter plot are (70, 170), (69, 165), (72, 175), (74, 180), (75, 185), (70, 175), (69, 168), (72, 173), (71, 178), (78, 190), (76, 188), and (69, 165).

Step 2: Draw the line that appears to best fit the data points. There should be about the same number of points above the line as below it. The line does not have to pass through any of the data points.



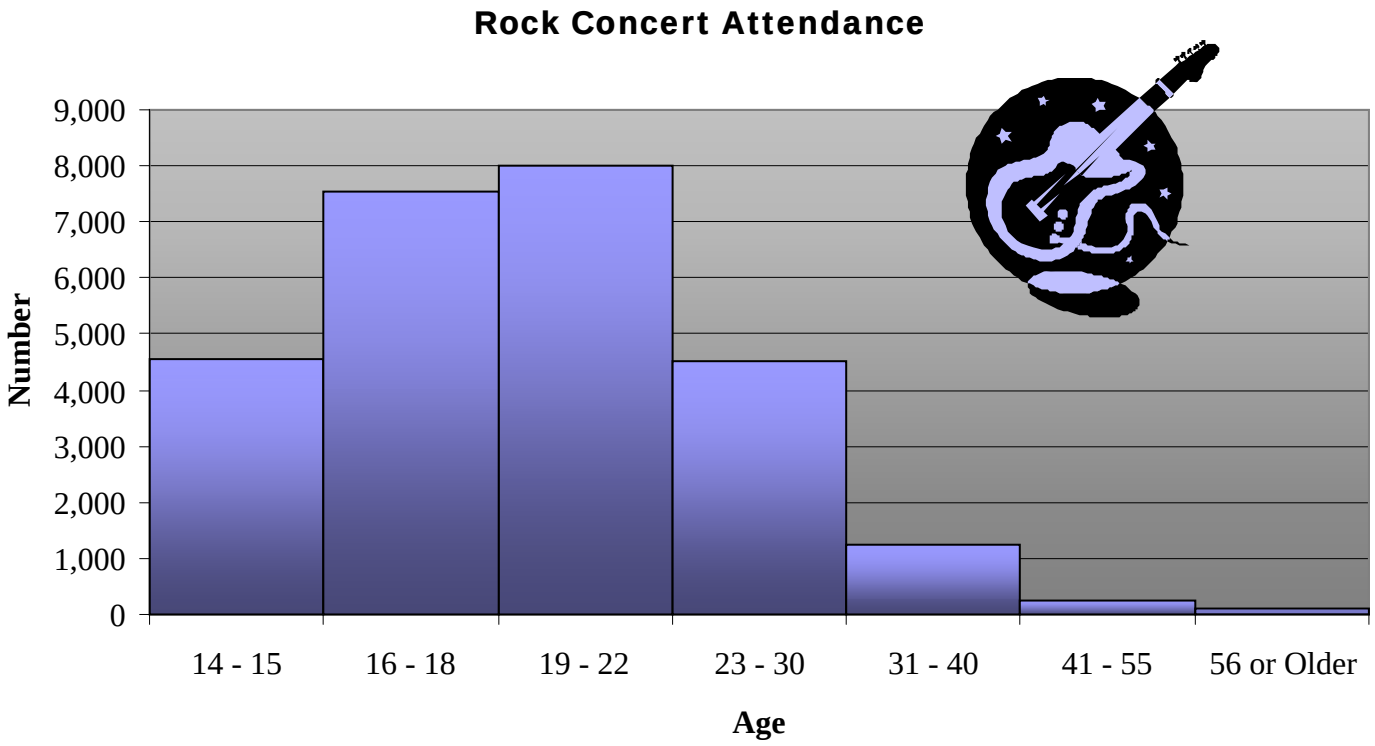
What kind of correlation exists between the data sets?

The scatter plot displays a positive correlation. The line of best fit slants upward and to the right indicating a positive correlation.

Histograms

Darin wanted to attend the rock concert being held at the local civic center. His mom agreed that he could go, but only if she went also. The data below depicts the ages of the persons attending the rock concert. It is also displayed in a histogram. Histograms are used to display ranges of data.

Ages	Frequency
14 – 15 (Young Teens)	4538
16 – 18 (Teens)	7536
19 – 22 (Young Adults)	7999
23 – 30 (Adults)	4520
31 – 40 (Middle-Aged Adults)	1236
41 – 55 (Older Adults)	233
56 or Older (Seniors)	120



What age group had the highest attendance? The 19-22 year olds

What age group had the lowest attendance? 56 or Older

In a histogram, a display of data which possesses a **symmetric** distribution is one in which the two “halves” of the data appear as mirror-images to one another.

A **skewed** (non-symmetric) distribution of data is a distribution in which the data appears at more diverse values. Distributions of data are skewed if one of the tails of a histogram (the part that stretches out from the peak) is longer than the other.

A "**skewed right**" distribution is one in which the right tail is longer.

A "**skewed left**" distribution is one in which the left tail is longer.

How would the distribution of the data of the rock concert attendance be described?

The data is “**skewed right**”. Most of the rock concert attendees were younger than 30 years old.

Sampling

When collecting data to make predictions, it is necessary to get an **unbiased** sample selection (small group) that will be representative of the population (whole group).

Suppose Rita wanted to determine the favorite after-school activity of the students in her class by surveying a sample of the entire class.

A **biased** sample would be a survey of the members of the computer club. These members have a common interest in computers so surveying them would probably reflect a lot of computer-related activities.

An **unbiased** sample would be a survey of every fifth person listed on the class roster in alphabetical order based on his/her last name.

Let's take a closer look at the types of biased and unbiased samples that are considered for surveys.

Types of biased samples:

convenience sample – A convenience sample is a sample that includes members of a population that are easily accessed.

voluntary response sample - A voluntary response sample is a sample that involves only those people that want to participate in the sampling.

Types of unbiased samples:

simple random sample - A simple random sample is a sample where each item or person in the population is as likely to be chosen as any other.

stratified random sample – A stratified random sample is a sample in which the population is divided into similar, non-overlapping groups.

systematic random sample – A systematic random sample is a sample in which the items or people are selected according to a specific time or item interval.

Examples: Identify the type of sample described.

1. A person employed by the local mall solicits shoppers to fill in a survey about new products by offering them a lottery ticket if they take the time to complete the survey.

This is a **voluntary response sample** (biased) because the participants are choosing to take the survey (most likely because they want the chance to win a prize with the lottery ticket).

2. Every person whose telephone number ends with a 48 is contacted to find out which presidential candidate he or she favors for the next election.

This is a **systematic random sample (unbiased)** because each person surveyed is selected from a list of most of the persons in the community and based upon the condition that his/her telephone number ends with a 48 (item interval).

3. The parents, grandparents, relatives, and friends who attend the school Christmas concert are surveyed and asked if they will support the next operating school tax levy.

This is a **convenience sample (biased)** because the people surveyed were the ones that were easily accessible because they attended the school's Christmas concert.

4. Persons, ages 30 to 39, are surveyed to see which car model they prefer to purchase.

This is a **stratified random sample (unbiased)** because a select age group, ages 30 to 39, were surveyed to see which type of car they prefer. Selections from other age groups may vary considerably but if the target group for sales is this age group, then the survey is unbiased.

Misleading Statistics

Statistics and graphs are very valuable tools for determining, processing, and evaluating data. However, some persons may misuse statistics purposely to promote an outcome they desire.

Example 1: Timothy purchased five raffle tickets that cost $1 each, based on the sales promotion that the top prize was $400 and the average award amount was $20. What the salesperson failed to mention to Timothy is that the rest of the prize awards were for $1.

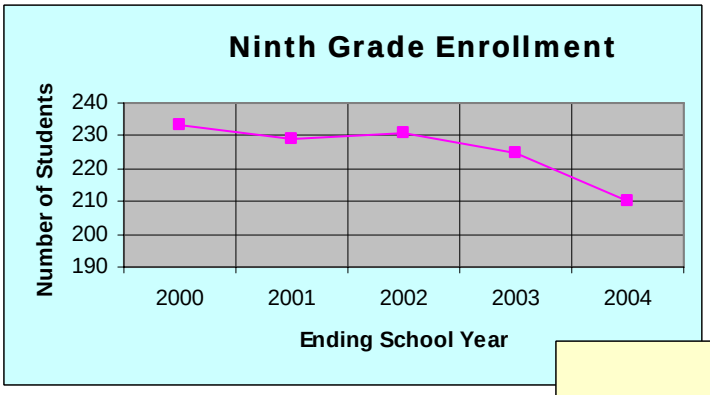
This promotion was mathematically correct, but if Timothy won, most likely his prize would be $1, not $400.

The Math:

1 prize of $400 = 400
20 prizes of $1 = 20
---
21 prizes = 420

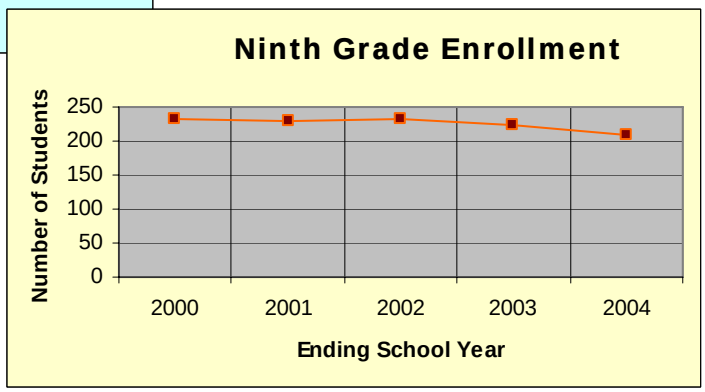
Average prize amount is $\frac{420}{21} = \$20$

Example 2: Graphs may be constructed in a way to emphasize certain data presented. Here are two graphs that show the enrollment of a ninth grade class over the school years 2000 through 2004.



In this graph, an overall decline in enrollment is shown with a sharp decline in the year 2004.

In this graph, the enrollment looks pretty steady with a small decline in the year 2004.



The different appearance of these two graphs was simply achieved by making changes in the beginning and ending values of the vertical axis and also by making changes in the amount of increments in the scale of the vertical axis.