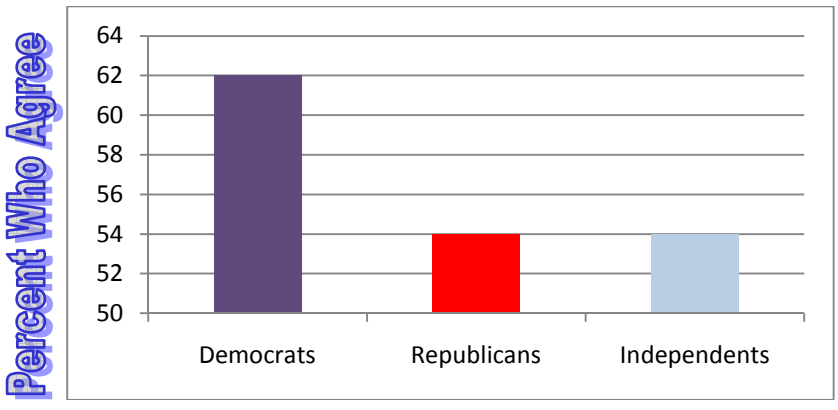
CHAPTER PROBLEM

Are the survey results presented in a way that is fair and objective?

At age 26, Terri Schiavo was married and was seeking to have a child when she collapsed from respiratory and cardiac arrest. Attempts to revive her were unsuccessful and she went into a coma. She was declared to be in a permanent vegetative state in which she appeared to be awake but unaware. She remained in that state for 15 years, unable to communicate or care for herself in any way. She was kept alive through the insertion of a feeding tube. There were intense debates about her situation, with some arguing that she should be allowed to die without the feeding tube, while others argued that her life should be preserved with the feeding tube and any other necessary means. After many legal battles, her feeding tube was removed, and Terri Schiavo died 13 days later at the age of 41. Although there were very different and strong opinions about Terri Schiavo's medical treatment, there was universal sympathy for her.

In the midst of the many debates about the removal of Terri Schiavo's feeding tube, there was a CNN/*USA Today*/Gallup poll in which respondents were asked this question: "Based on what you have heard or read about the case, do you agree with the court's decision to have the feeding tube removed?" The survey was conducted by telephone and there were 909 responses from adults in the United States. Respondents were also asked about their political party affiliations, and a bar graph similar to the one below was placed on the CNN website. This figure shows the poll results broken down by political party. Based on this figure, it appears that responses by Democrats were substantially different from responses by Republicans and Independents.

We will not address the human issues related to the removal of the feeding tube, although it raises important questions that everyone should carefully consider. Instead, we will focus on the graph. Our understanding of graphs and the information they convey will help us answer this question: Does the graph fairly represent the survey results?



MATH 103 CHAPTER 2 HOMEWORK			
2.1	NA	**2.4**	1-4, 9-14, 19-21, 23, 25
2.2	1-5, 7, 9, 11, 13, 16, 17, 20, 25, 29	**2.5**	1, 3, 4, 6, 7, 8, 10
2.3	1-8, 11, 17, 19		

2.1 REVIEW AND PREVIEW

Chapter 1 discussed _____ thinking and methods for _____ data and identifying _____ of data. We also discussed consideration of the _____ of the data, the _____ of the data and the _____ method. Samples of data are often _____; in order to analyze such large data sets, we must _____, _____, and _____ data in a convenient and meaningful form.

In this chapter, we are mainly concerned with the _____ of the data set.

CHARACTERISTICS OF DATA

1. **Center**: A representative or average value that indicates where the
2. _____ of the data set is located.
3. **Variation**: A measure of the amount that data values _____.

4. **Distribution**: The nature or shape of the _____ of the data over the _____ of values (such as bell-shaped, uniform, or skewed).
5. **Outliers**: Sample values that lie very far away from the vast _____ of the other sample values.
6. **Time**: Changing characteristics of the data over time.

2.2 FREQUENCY DISTRIBUTIONS

Key Concept…

When working with large data sets, it is often helpful to _____ and _____ the data by constructing a table called a _____ distribution. A frequency distribution helps us understand the nature of the distribution of a data set.

DEFINITION

A **frequency distribution (aka frequency table)** shows how a data set is _____ among all of several categories (or classes) by listing all of the _____ along with the number of data _____ in each of the categories.

Example 1: Let's construct our own frequency distribution which summarizes the height distribution in our class.

Height of Students in Ms. Gracey's Class

HEIGHT	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74
FREQUENCY								

DEFINITION

Lower class limits are the _____ numbers that can belong to the different _____.

Upper class limits are the _____ numbers that can belong to the different _____.

Class boundaries are the numbers used to _____ the classes, but without the gaps created by class limits.

Class midpoints are the values in the _____ of the classes. Each class midpoint is found by adding the lower class limit to the upper class limit and dividing the sum by 2.

Class width is the difference between two consecutive lower class limits or two consecutive lower class boundaries.

35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74

PROCEDURE FOR CONSTRUCTING A FREQUENCY DISTRIBUTION

We construct frequency distributions so that (1) large data sets can be

_____, (2) we can analyze the _____

of the data, and (3) we have a basis for constructing _____ (such as histograms).

1. Determine the number of _____. The number of classes should be between 5 and 20, and the number you select might be affect by the convenience of using large numbers.
2. Calculate the class width.

$$\text{class width} \approx \frac{(\qquad\qquad) - (\qquad\qquad)}{\text{number of classes}}$$

3. Choose either the minimum data value or a convenient value _____ the minimum data value as the first lower class limits.
4. Using the first lower class limit and the class width, list the other lower class limits. (Add the class width to the _____ lower class limit to get the second lower class limit. Add the class width to the _____ lower class limit to get the third lower class limit, and so on).
5. List the lower class limits in a vertical column and then enter the upper class limits.
6. Take each individual data value and put a tally mark in the appropriate class. _____ the tally marks to find the total frequency for each class.

RELATIVE FREQUENCY DISTRIBUTION

In a relative frequency distribution, the frequency of a class is replaced with a relative frequency (aka a proportion) or a percentage frequency. The sum of the

relative frequencies in a relative frequency distribution must be close to _____

or _____.

$$\text{relative frequency} = \frac{\qquad\qquad}{\qquad\qquad}$$

$$\text{percentage frequency} = \frac{\qquad\qquad}{\qquad\qquad} \times 100$$

CUMULATIVE FREQUENCY DISTRIBUTION

The cumulative frequency for a class is the _____ of the

_____ for that class and all

_____ classes.

CRITICAL THINKING: INTERPRETING FREQUENCY DISTRIBUTION

In statistics, we are interested in the _____ of the

data, and in particular, whether the data have a _____ distribution.

NORMAL DISTRIBUTION

1. The _____ start low, then increase to one or two high frequencies, then decrease to a low frequency.

2. The distribution is approximately _____, with frequencies preceding the maximum being roughly a mirror image of those that follow the maximum.

GAPS

The presence of gaps can show that we have data from two or more different

_____. BE CAREFUL—the converse is not necessarily true!

Example 2: Consider the frequency distribution below.

Tar (mg) in filtered cigarettes	Frequency
2-5	2
6-9	2
10-13	6
14-17	15

a. Identify the class width

b. Identify the class midpoints

c. Identify the class boundaries

d. Using a strict interpretation of the relative criteria, does the frequency distribution appear to have a normal distribution?

e. If the criteria are interpreted very loosely, does the frequency distribution appear to have a normal distribution?

Example 3: Listed below are amounts of strontium-90 (in millibecquerels) in a simple random sample of baby teeth obtained from Pennsylvania residents born after 1979. Construct a frequency distribution with eight classes. Begin with a lower class limit of 110, and use a class width of ten.

155 142 149 130 151 163 151 142 156 133 138 161 128 144 172 137 151
166 147 163 145 116 136 158 114 165 169 145 150 150 158 151 145 152
140 170 129 188 156

2.3 HISTOGRAMS

Key Concept…
In this section we discuss a visual tool called a histogram, and its significance in representing and analyzing data.

DEFINITION

A **histogram** is a graph consisting of bars of _____ width drawn adjacent to each other without _____. The horizontal scale represents _____ of quantitative data values and the vertical scale represents _____.

HORIZONTAL SCALE: Use class _____ or class _____

VERTICAL SCALE: Use class _____

A **relative frequency histogram** has the same shape and horizontal scale as a histogram, but the vertical scale is marked with _____ frequencies (as _____ or _____) instead of actual frequencies.

CRITICAL THINKING: INTERPRETING HISTOGRAMS

We _____the histogram to see what we can learn about

C_____

V_____

D_____

O_____

T_____

Example 4: Use the frequency distribution from example 3 to construct a histogram.

2.4 STATISTICAL GRAPHICS

Key Concept…
In this section we discuss types of statistical graphs other than

_____. Our objective is to identify a

_____ graph for representing a

_____ set.

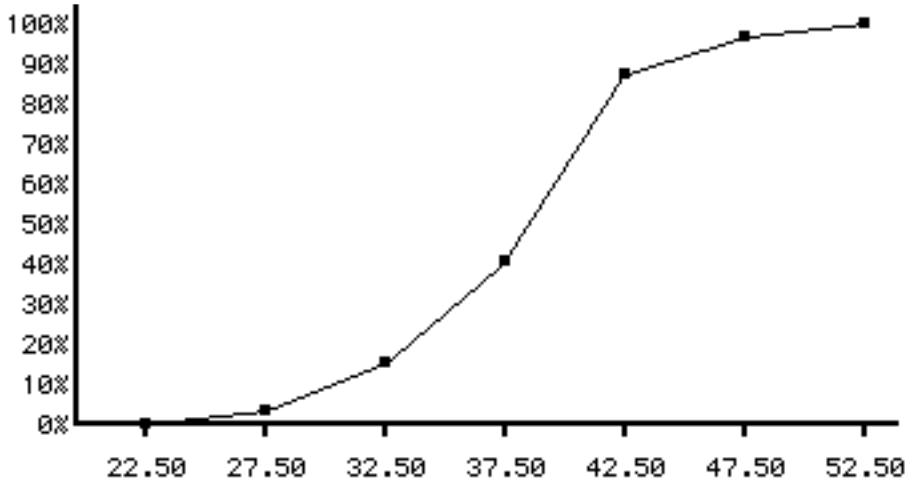
FREQUENCY POLYGON

A **frequency polygon** uses line segments connected to points directly above class

_____ values.

A **relative frequency polygon** uses relative frequencies for the

_____ scale.



OGIVE

An **ogive (pronounced "oh-jive")** involves _____ frequencies. Ogives are useful for determining the number of values below some particular value. An ogive is a _____ graph that depicts cumulative frequencies. An ogive uses class boundaries along the horizontal scale, and _____ frequencies along the vertical scale.

For example, if you saved $300 in both January and April and $100 in each of February, March, May, and June, an ogive would look like Figure 1 .

Figure 1 Ogive of accumulated savings for one year.

DOTPLOTS

A **dotplot** consists of a graph in which each data value is plotted as a _____ or _____ along a scale of values.

Dots representing equal values are _____.

STEMPLOTS

A **stemplot (aka stem-and-leaf plot)** represents

_____ data by separating each value into

two parts: the _____ and the _____.

Example 1: Listed below are amounts of strontium-90 (in millibecquerels) in a simple random sample of baby teeth obtained from Pennsylvania residents born after 1979.

155 142 149 130 151 163 151 142 156 133 138 161 128 144 172 137 151
166 147 163 145 116 136 158 114 165 169 145 150 150 158 151 145 152
140 170 129 188 156

a. Construct a dotplot of the amounts of Strontium-90

i. What does the dotplot suggest about the distribution?

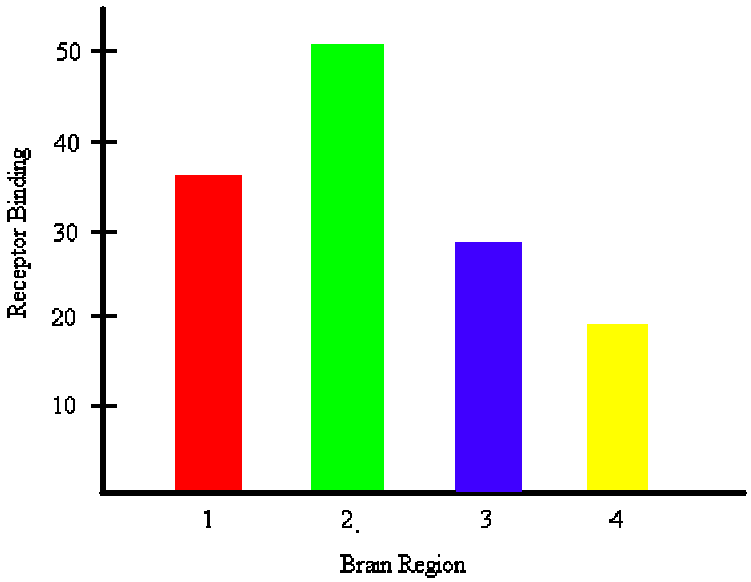
b. Construct a stemplot of the amounts of Strontium-90

i. What does the stemplot suggest about the distribution?

BAR GRAPH

A **bar graph** uses bars of _____ width to show frequencies of categories of _____ data. The vertical scale represents _____ or _____ frequencies. The horizontal scale identifies the different _____ of qualitative data. The bars may or may not be separated by small gaps.

A **multiple bar graph** has two or more sets of bars, and is used to compare two or more _____ sets.



PARETO CHARTS

A **Pareto chart** is a bar graph for _____ data, with added stipulation that the bars are arranged in descending order according to

_____. The vertical scale represents

_____ or _____ frequencies. The horizontal scale identifies the different categories of

_____ data.

PIE CHARTS

A **pie chart** is a graph that depicts _____ data as

slices of a _____, in which the size of each slice is proportional to the frequency count for each category.

Example 2: Chief financial officers of U.S. companies were surveyed about areas in which job applicants make mistakes. Here are the areas and the frequency of

responses: interview (452); résumé (297); cover letter (141); reference checks (143); interview follow-up (113); screening call (85).

a. Construct a pie chart representing the given data.

b. Construct a Pareto chart of the data.

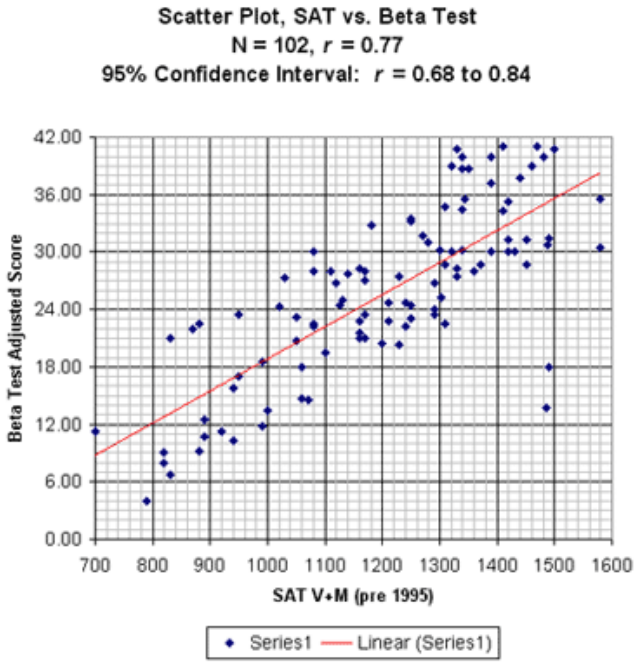
c. Which graph is more effective in showing the importance of the mistakes made by job applicants?

SCATTERPLOTS

A **scatterplot (aka scatter diagram)** is a plot of ordered pair

_____ data with a horizontal *x*-axis and a vertical *y*-axis. The horizontal axis is used for the first (*x*) variable, and the vertical axis is used for the second variable. The pattern of the plotted points is often helpful

in determining whether there is a _____ between the two variables.

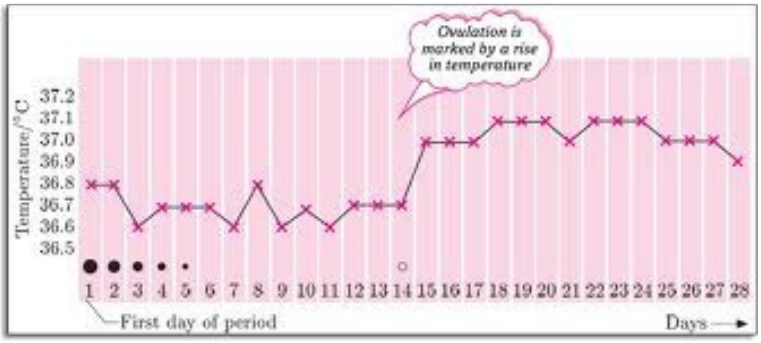


TIME-SERIES GRAPH

A **time-series graph** is a graph of *time-series data,* which are

_____ data that have been collected at

different points in _____.



2.5 CRITICAL THINKING: BAD GRAPHS

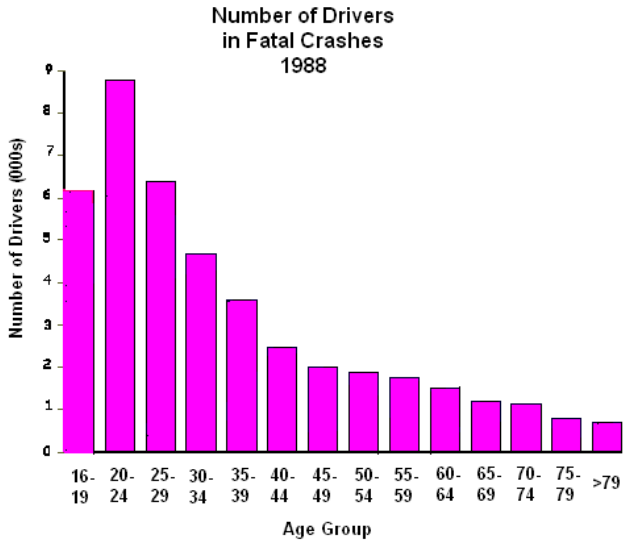
Key Concept…

Some graphs are bad because they are technically correct, but _____. In this section we will learn about two of the most common types of bad graphs.

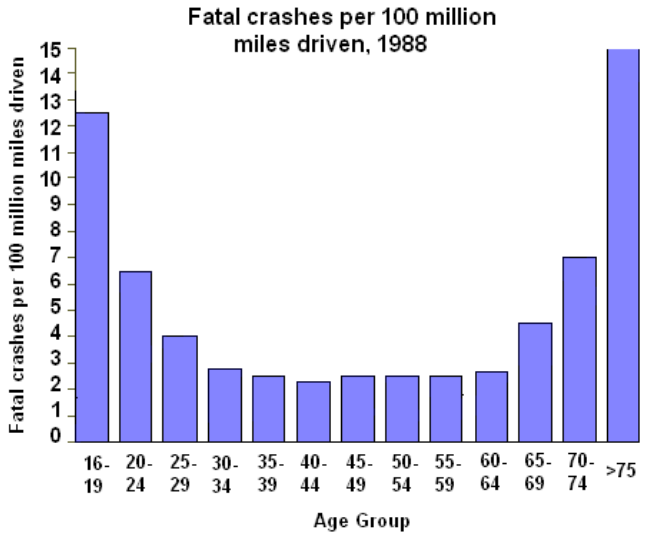
Nonzero axis

Some graphs are misleading because one or both of the _____ begin at some value other than _____, so the differences are _____.

1. The following statistics suggest that 16-year-olds are safer drivers than people in their twenties, and that octogenarians are very safe. Is this true?



Graph is based on data from this study: Williams, Allan F., Ph.D., and Oliver Carston, Ph.D., "Driver Age and Crash Involvement," Am J Public Health 1989; 79: 326-327.



Graph is based on data from this study: Williams, Allan F., Ph.D., and Oliver Carston, Ph.D., "Driver Age and Crash Involvement," Am J Public Health 1989; 79: 326-327.

Solution: No. As the following graph shows, the reason 16-year-old and octogenarians appear to be safe drivers is that they don't drive nearly as much as people in other age groups.

Pictographs

Drawings of objects, often called pictographs, are often misleading.