2 | DISPLAYING DATA

Figure 2.1 Photo by [Myriam Jessier](#) on [Unsplash](#)

Introduction

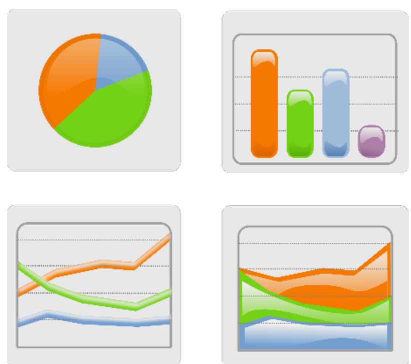
Chapter Objectives

By the end of this chapter, the student should be able to:

- Create and interpret frequency distributions.
- Display data graphically and interpret graphs for Qualitative Data (Bar, Pareto, Pie)
- Display data graphically and interpret graphs for Quantitative Data (Histogram, Frequency Polygon, and Time Series)

Once you have collected data, what will you do with it? Data can be described and presented in many different formats. For example, suppose you are interested in buying a house in a particular area. You may have no clue about the house prices, so you might ask your real estate agent to give you a sample data set of prices. Looking at all the prices in the sample often is overwhelming. A better way might be to look at the median price and the variation of prices. The median and variation are just two ways that you will learn to describe data. Your agent might also provide you with a graph of the data. In this chapter, you will study numerical and graphical ways to describe and display your data. This area of statistics is called "Descriptive Statistics." You will learn how to calculate, and even more importantly, how to interpret these measurements and graphs.

A statistical graph is a tool that helps you learn about the shape or distribution of a sample or a population. A graph can be a more effective way of presenting data than a mass of numbers because we can see where data clusters and where there are only a few data values. Newspapers and the Internet use graphs to show trends and to enable readers to compare facts and figures quickly. Statisticians often graph data first to get a picture of the data. Then, more formal tools may be applied.



https://openclipart.org/detail/19980/graphs

Some of the types of graphs that are used to summarize and organize data are the dot plot, the bar graph, the histogram, the stem-and-leaf plot, the frequency polygon (a type of broken line graph), the pie chart, and the box plot. In this chapter, we will briefly look at stem-and-leaf plots, line graphs, and bar graphs, as well as frequency polygons, and time series graphs. Our emphasis will be on histograms and box plots.

NOTE: This book contains instructions for constructing a histogram and a box plot for the TI-83+ and TI-84 calculators. The Texas Instruments (TI) website (http://education.ti.com/educationportal/sites/US/sectionHome/ support.html) provides additional instructions for using these calculators.

2.1 | Frequency Distributions

Once you have a set of data, you will need to organize it so that you can analyze how frequently each datum occurs in the set. This is done by creating a table of frequencies which is known as a **frequency distribution**. There are three frequency distributions to choose from, depending on if the data is categorical or quantitative. When it is quantitative data, there are two frequency distributions to choose from. When calculating the frequency, you may need to round your answers so that they are as precise as possible.

Answers and Rounding Off

A simple way to round off answers is to carry your final answer one more decimal place than was present in the original data. Round off only the final answer. Do not round off any intermediate results, if possible. If it becomes necessary to round off intermediate results, carry them to at least twice as many decimal places as the final answer. For example, the average of the three quiz scores four, six, and nine is 6.3, rounded off to the nearest tenth, because the data are whole numbers. Most answers will be rounded off in this manner.

Ungrouped Frequency Distribution

This first type of frequency distribution is for quantitative data sets. It is best when the range of the data is less than 10 units. **Range** is the difference between the largest data value and the smallest data value. For example, twenty students were asked how many hours they worked per day. Their responses, in hours, are as follows:

5; 6; 3; 3; 2; 4; 8; 5; 2; 3; 5; 6; 5; 4; 4; 3; 5; 2; 5; 3.

Since the range is 6, we will keep each data value separate and not group them together. To create an ungrouped frequency distribution is a simple task. Place the data values from smallest to the largest without skipping any values on the first column. Place the **frequency**, the count of each data value, in the corresponding row of the second column.

Table 2.1 lists the different data values in ascending order and their frequencies. Notice all the data values are listed including 7 which is not listed on the original data set.

DATA VALUES	FREQUENCY (f)
2	3
3	5
4	3
5	6
6	2
7	0
8	1

Table 2.1 Frequency distribution of Student Work Hours

A **frequency** is the number of times a value of the data occurs. According to **Table 2.1**, there are three students who work two hours, five students who work three hours, and so on. The sum of the values in the frequency column, 20, represents the total number of students included in the sample, known as **sample size** (n).

A **relative frequency** is the ratio (fraction or proportion) of the number of times a value of the data occurs in the set of all outcomes to the total number of outcomes. To find the relative frequencies, divide each frequency by the total number of students in the sample, **n**, in this case, 20. Relative frequencies can be written as fractions, decimals, or percent.

The following table is known as a relative frequency distribution because frequencies are translated as percent. To find the percent of the first row, take the frequency and divide by the sample size, n. 3/20 = 15%.

DATA VALUE	FREQUENCY (f)	RELATIVE FREQUENCY
2	3	15%
3	5	25%
4	3	15%
5	6	30%
6	2	10%
7	0	0%
8	1	5%

Table 2.2 Frequency distribution of Student Work Hours with Relative Frequencies

Cumulative relative frequency is the accumulation of the previous relative frequencies. To find the cumulative relative frequencies, add all the previous relative frequencies to the relative frequency for the current row, as shown in **Table 2.3**. The third row under cumulative relative frequency has the number .55 which was found by adding the previous relative frequencies (.15 + .25 + .15). This represents that 55% of the data has the value of 4 or under.

DATA VALUE	FREQUENCY (f)	RELATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY
2	3	15%	.15
3	5	25%	.40
4	3	15%	.55
5	6	30%	.85
6	2	10%	.95
7	0	0%	.95
8	1	5%	1

Table 2.3 Frequency distribution of Student Work Hours with Relative Frequencies

Grouped Frequency Distribution

This second type of frequency distribution is also used when there is quantitative data. However, it is used when the range is large and the data values need to be grouped together. For example, 28 students were asked how many hours they worked per week. Their responses, in hours, are as follows:

15; 26; 13; 33; 22; 14; 27; 15; 32; 23; 5; 26; 25; 14; 34; 13; 15; 22; 15; 28; 10; 18; 21; 24; 20; 18; 34; 20;

Here there are too many different data values to list them separately as in the ungrouped frequency distribution. Notice the range is 29 (highest – lowest = 34 – 5). Therefore we need to construct a grouped frequency distribution (**Table 2.4**) to group data values into classes. A **class** is an interval where the lowest value of the interval is known as the **lower limit** and the highest value of the interval is known as the **upper limit.**

Guidelines for classes:

- There should be between 5 and 20 classes
- Classes must be mutually exclusive (no overlap of data values)
- Classes must be all inclusive and continuous
- Classes must be equal in width

Constructing a Grouped Frequency Distribution:

1. Find Range (highest data value – lowest data value)

2. Determine the number of classes (usually the minimum is 5 classes and a maximum of 20 classes)

3. Find Class Width = $\dfrac{\text{Range}}{\text{\# of classes}}$ NOTE: Round up

4. Choose first lower limit (usually the lowest data value)

5. Create the other lower limits of the classes by adding the class width to the previous lower limit

6. Create the upper limits by not overlapping the limits

Example 2.1

Twenty-eight students were asked how many hours they worked per week. Their responses, in hours, are as follows: 15; 26; 13; 33; 22; 14; 27; 15; 32; 23; 5; 26; 25; 14; 34; 13; 15; 22; 15; 28; 10; 18; 21; 24; 20; 18; 34; 20; construct a grouped frequency distribution using 5 classes.

5; 10; 13; 13; 14; 14; 15; 15; 15; 15; 18; 18; 20; 20; 21; 22; 22; 23; 24; 25; 26; 26; 27; 28; 32; 33; 34;

1. Range = 34 – 5 = 29

2. Use 5 classes (given in **Exercise 2.1**)

3. Class Width = 29/5 = 5.8 round up to 6

4. First lower limit will be 5 which is the minimum data value

5. The other lower limits will be 11, 17, 23, 29 by adding the class width of 6 to the previous lower limit

6. The first upper limit will be 10 since the next class begins at 11. Using class width again, the other upper limits are 16, 22, 28, 34

Solution 2.1

CLASSES	FREQUENCY (f)
5 – 10	2
11 – 16	8
17 – 22	7
23 – 28	7
29 – 34	4

Frequency distribution of Student Work Hours.

Try It Σ

2.1 The grade point averages for 40 students are listed below.

2.0	3.2	1.8	2.9	0.9	4.0	3.3	2.9	3.6	0.8
3.1	2.4	2.4	2.3	1.6	1.6	4.0	3.1	3.2	1.8
2.2	2.2	1.7	0.5	3.6	3.4	1.9	2.0	3.0	1.1
3.0	4.0	4.0	2.1	1.9	1.1	0.5	3.2	3.0	2.2

Construct a frequency distribution, a relative frequency distribution, and a cumulative frequency distribution using eight classes. Include the midpoints of the classes.

Solution 'Try It' 2.1:

Range = 4 - .5 = 3.5

Class width = 3.5/8 = .4375 Round up to .5 since the data values are in tenths (one decimal spot) then we round the class width to tenths.

CLASSES	FREQUENCY (f)
.5 - .9	4
1 – 1.4	2
1.5 – 1.9	7
2 – 2.4	9
2.5 – 2.9	2
3-3.4	10
3.5-3.9	2
4-4.4	4

Table 2.5 Frequency distribution of GPA

Frequency Distribution for Qualitative Data

Qualitative data is non-numeric data. Therefore, when creating a table of data values, place the category in the first column and the count of each data value, frequency, in the second column. For example, twenty students are asked their blood type. Their responses are as follows:

A; B; O; A; AB; O; O; A; O; B; A; A; A; O; O; O; B; O; AB; B

Table 2.5 lists the different data values and their frequencies.

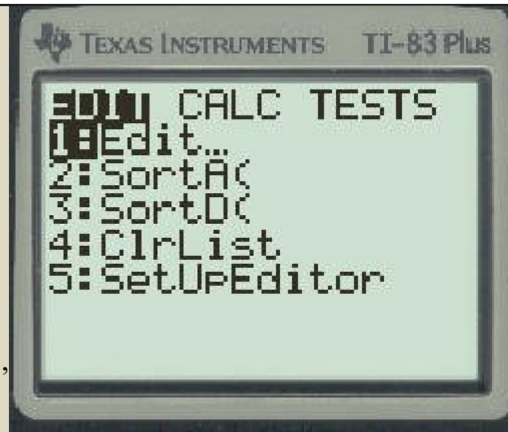
DATA VALUE	FREQUENCY (f)
A	6
AB	2
B	4
O	8

Table 2.5 Frequency distribution of Student Blood Types

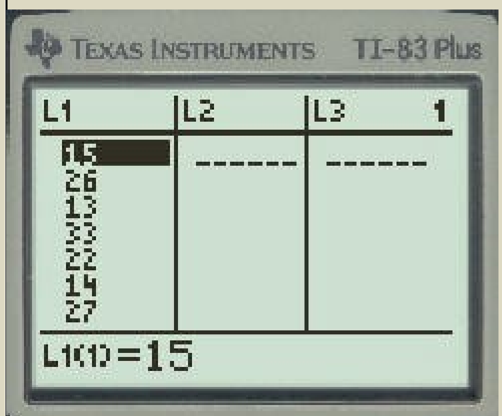
Frequency distributions (tables) are a good way of organizing and displaying data. But graphs can be even more helpful in understanding the data

The TI-83/TI-84 Calculator can sort the data values

Press STAT
Choose EDIT



Once you have enter the data into L1, 2nd QUIT
Press STAT



Choose #2 SortA(
Press 2nd L1

To view the sorted list, return to STAT, EDIT

2.2 | Graphs for Qualitative Data

There are no strict rules concerning which graphs to use. However, scale of the bars or slices should be accurate. For example, if a category is 32% then the slice for the category should not be smaller than a quarter of the circle or bigger than half the circle. Three graphs that are used to display qualitative data are pie charts, bar graphs, and Pareto charts.
In a **pie chart**, categories of data are represented by wedges in a circle and are proportional in size to the percent of individuals in each category. When creating a pie chart, each slice should be labeled with the category name and the relative frequency (percent).

In a **bar graph**, the length of the bar for each category is proportional to the number or percent of individuals in each category. Bars may be vertical or horizontal. For vertical bars, the categories are on the x-axis and frequency or relative frequency are on the y-axis.

In a **Pareto chart** consists of bars that are sorted into order by category size (largest to smallest). Pareto charts are typically used with nominal data. If a Pareto chart were created from ordinal or quantitative data, the values on the x-axis might seem out of order after the bars were rearranged from largest to smallest.

Pie Graphs:

Example 2.2

Below is a table comparing the number of part-time and full-time students at De Anza College and Foothill College enrolled for the spring 2010 quarter. The tables display counts (frequencies) and percentages or proportions (relative frequencies).

De Anza College			**Foothill College**		
	frequency	relative frequency		frequency	relative frequency
Full-time	9,200	40.9%	Full-time	4,059	28.6%
Part-time	13,296	59.1%	Part-time	10,124	71.4%
Total	22,496	100%	Total	14,183	100%

Table 2.6 Fall Term 2007 (Census day)

Solution 2.2

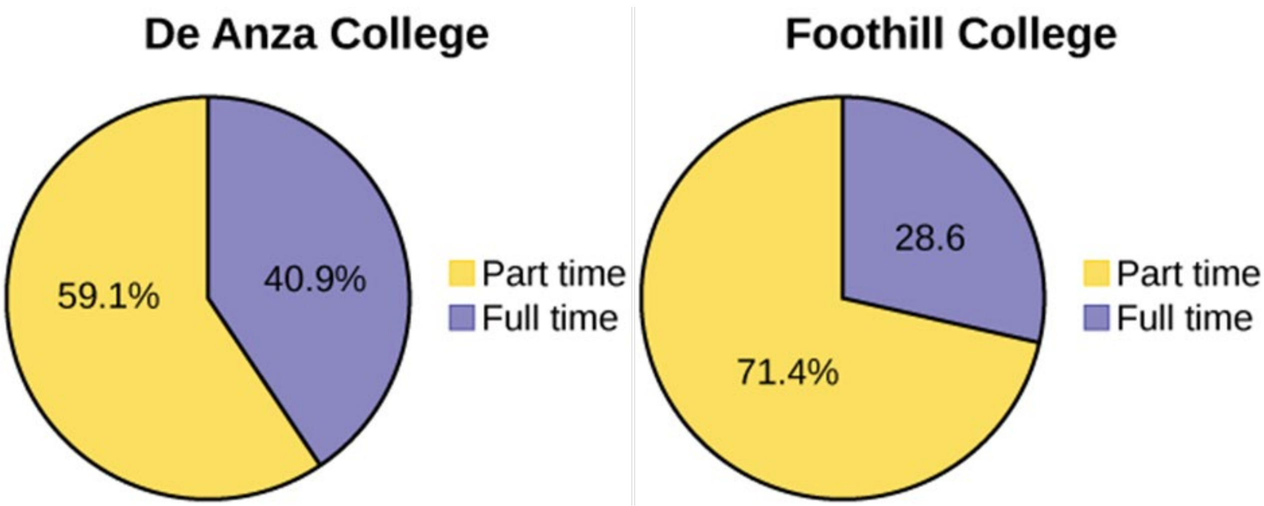
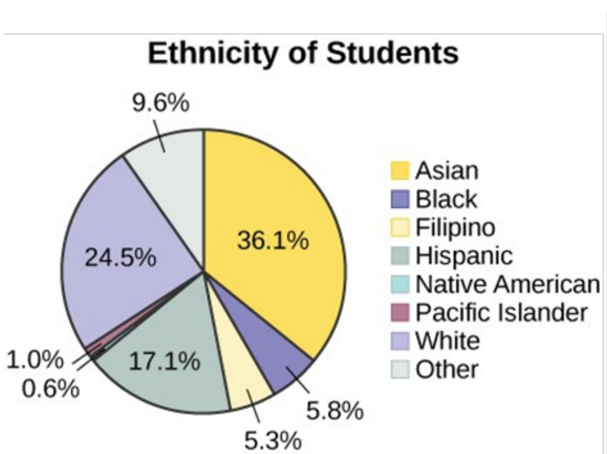


Figure 2.1 Pie Graphs

Pie Charts: No Missing Data

The following pie charts have the "Other/Unknown" category included (since the percentages must add to 100%). The chart in **Figure 2.7b** is organized by the size of each wedge, which makes it a more visually informative graph than the unsorted, alphabetical graph in **Figure 2.7a**.



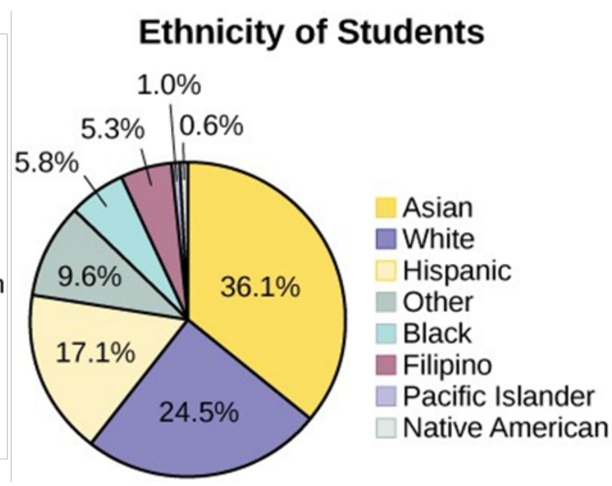


Figure 2.7 a Figure 2.7 b

Bar Graphs:

Below is the same data from Example 2.2 displayed as a bar graph.

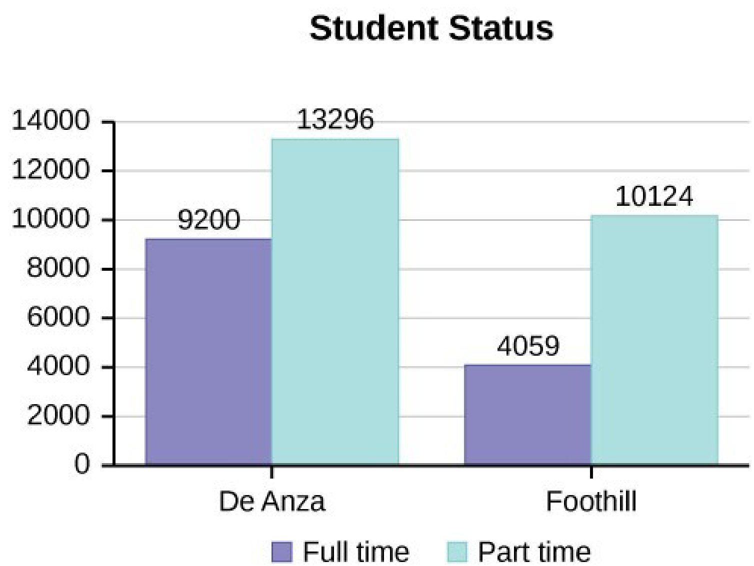


Figure 2.2 Bar Graph

Look at **Figure 2.1** and **Figure 2.2** and determine which graph (pie or bar) you think displays the comparisons better. It is a good idea to look at a variety of graphs to see which is the most helpful in displaying the data. We might make different choices of what we think is the "best" graph depending on the data and the context. Our choice also depends on what we are using the data for.

Percentages That Add to More (or Less) Than 100%

Sometimes percentages add up to be more than 100% (or less than 100%). In the graph, the percentages add to more than100% because students can be in more than one category. A **bar graph** is appropriate to compare the relative size of the categories. A pie chart cannot be used. It also could not be used if the percentages added to less than 100%.

Characteristic/Category	Percent
Full-Time Students	40.9%
Students who intend to transfer to a 4-year educational institution	48.6%
Students under age 25	61%
Total	150.5%

Table 2.7 De Anza College Spring 2010

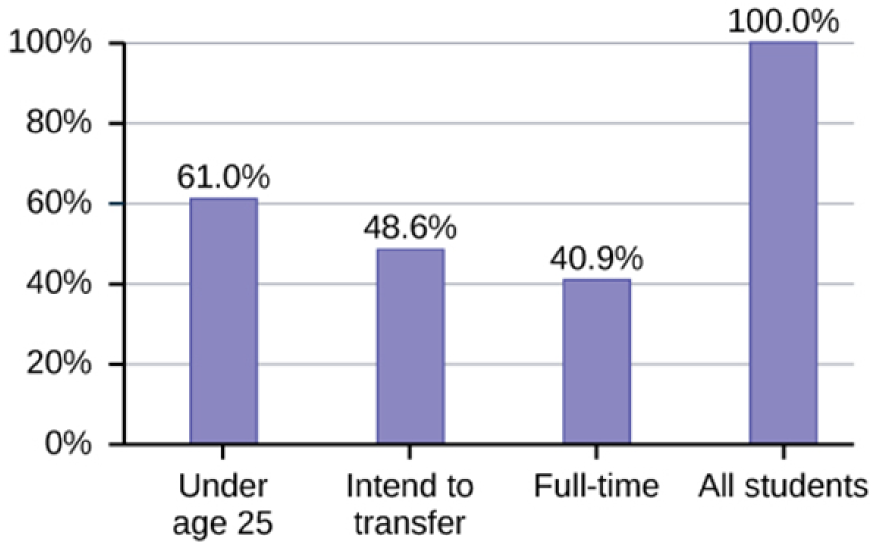


Figure 2.3 Bar Graph

Omitting Categories/Missing Data

The table displays Ethnicity of Students but is missing the "Other/Unknown" category. This category contains people who did not feel they fit into any of the ethnicity categories or declined to respond. Notice that the frequencies do not add up to the total number of students. In this situation, create a **bar graph** and not a pie chart.

	Frequency	**Percent**
Asian	8,794	36.1%
Black	1,412	5.8%
Filipino	1,298	5.3%
Hispanic	4,180	17.1%
Native American	146	0.6%
Pacific Islander	236	1.0%
White	5,978	24.5%
TOTAL	22,044 out of 24,382	90.4% out of 100%

Table 2.8 Ethnicity of Students at De Anza College Fall Term 2007 (Census Day)

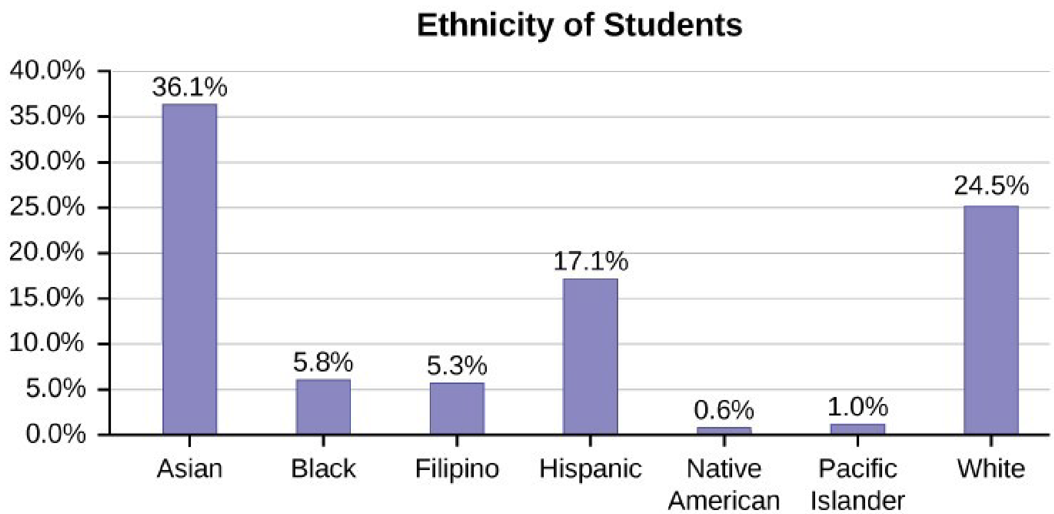


Figure 2.4 Bar Graph

The following graph is the same as the previous graph but the "Other/Unknown" percent (9.6%) has been included. The "Other/Unknown" category is large compared to some of the other categories (Native American, 0.6%, Pacific Islander 1.0%). This is important to know when we think about what the data are telling us. This particular bar graph in **Figure 2.5** can be difficult to understand visually.

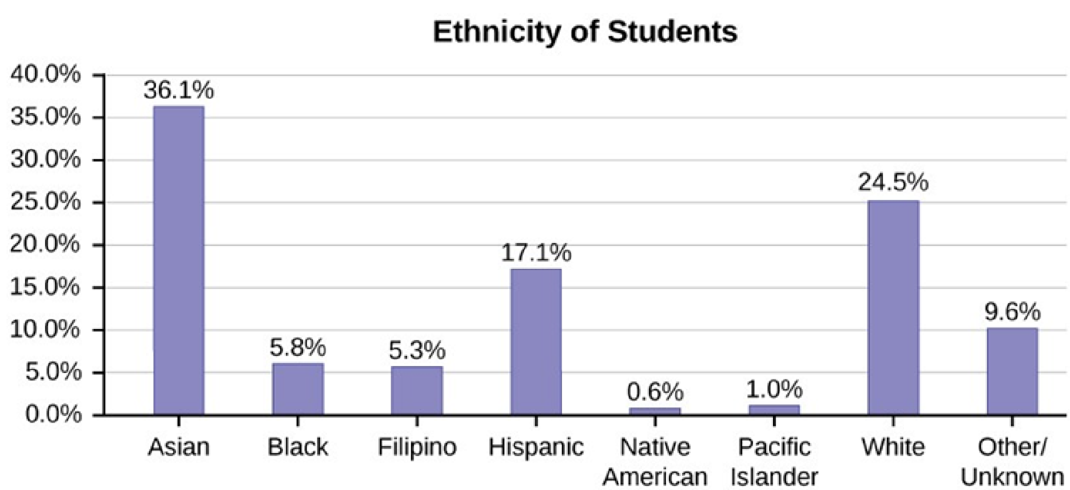


Figure 2.5 Bar Graph with Other/Unknown Category

Pareto Charts:

As the above data may be more difficult to interpret in a bar graph, a Pareto chart is easier to read and interpret. The graph in **Figure 2.6** is a Pareto chart.

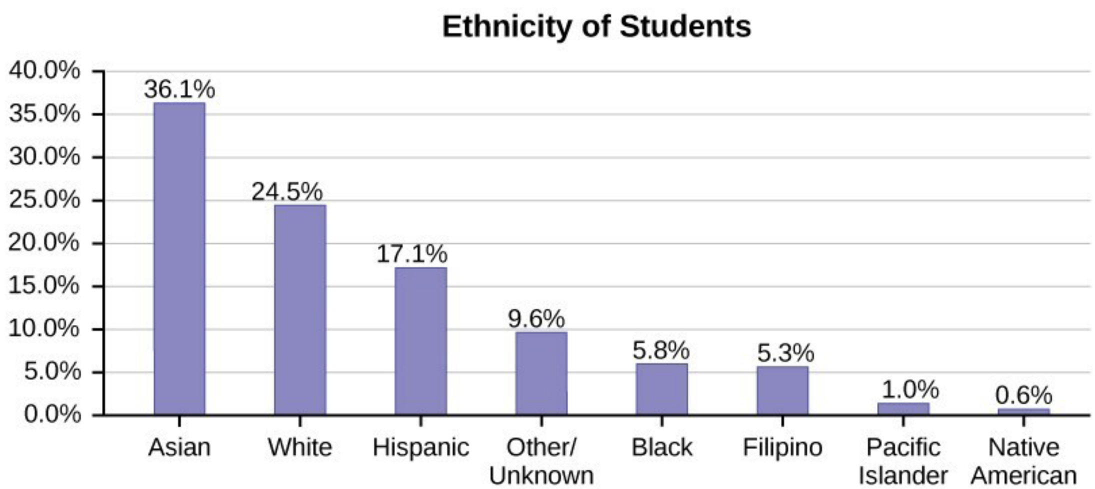


Figure 2.6 Pareto Chart

Example 2.3

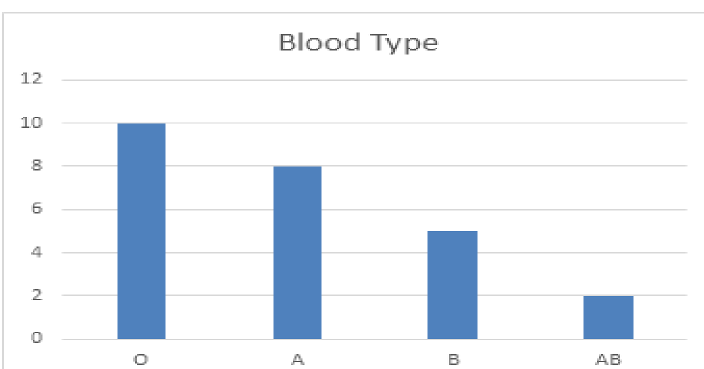
Twenty-five students are asked their blood type. Their responses are as follows:
A; B; O; A; AB; O; O; A; O; B; A; A; A; O; O; O; B; O; AB; B, O, B, O, A, A

DATA VALUE	FREQUENCY (f)
A	8
AB	2
B	5
O	10

Create a Pareto Chart.

Solution 2.3

Blood type O has the largest frequency therefore it should be placed first on the x-axis and blood type A follows.



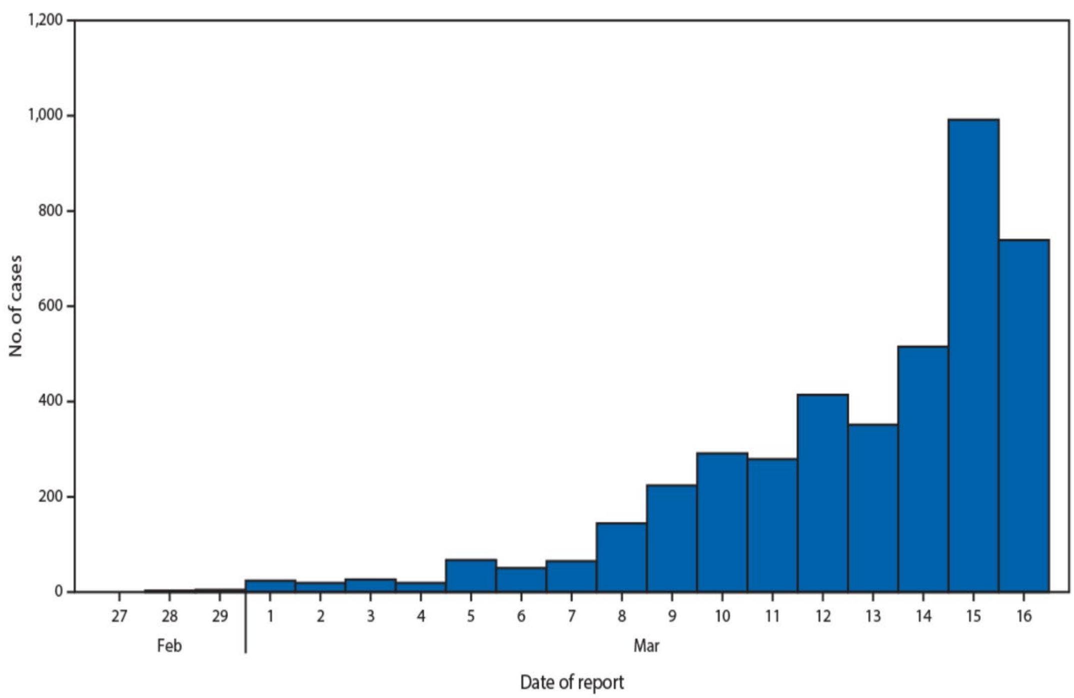
2.3 | Graphs for Quantitative Data

Histogram:

For most of the work you do in this book, you will use a histogram to display the data. One advantage of a histogram is that it can readily display large data sets. A *rule of thumb* is to use a histogram when the data set consists of 100 values or more.

A histogram consists of contiguous (adjoining) bars. It has both a horizontal axis and a vertical axis. The horizontal axis is labeled with what the data represents (for instance, distance from your home to school). Horizontal axis uses the class boundaries. The vertical axis is labeled either frequency or relative frequency (percent or probability). The graph will have the same shape with either label. The histogram can give you the shape of the data, the center, and the spread of the data.

Below is a histogram of the number of new coronavirus disease (Covid-19) cases reported daily*† (N = 4,226) – United States, February 12 – March 16, 2020 from the CDC (https://www.cdc.gov/mmwr/volumes/69/wr/mm6912e2.htm)



* Includes both COVID-19 cases confirmed by state or local public health laboratories, as well as those testing positive at the state or local public health laboratories and confirmed at CDC.

† Cases identified before February 28 were aggregated and reported during March 1–3.

Notice that there are no gaps between the bars and that date is in the middle of each bar. The class boundaries are where the bars meet. The horizontal axis may also be labeled by the class boundaries.

The **relative frequency** is equal to the frequency for an observed value of the data divided by the total number of data values in the sample. (Remember, frequency is defined as the number of times an answer occurs.)

- f = frequency
- n = total number of data values (or the sum of the individual frequencies)
- RF = relative frequency

$$RF = \frac{f}{n}$$

For example, if three students in Mr. Ahab's English class of 40 students received from 90% to 100%, then, f = 3, n = 40, and RF = 3/40 = .075 = 7.5% of the students received 90–100%.

To construct a histogram:
1. Create class boundaries on the grouped frequency distribution. Choose a starting point for the first interval to be less than the smallest data value. A convenient starting point is a lower value carried out to one more decimal place than the value with the most decimal places. For example, if the value with the most decimal places is 6.1 and this is the smallest value, a convenient starting point is 6.05 (6.1 – 0.05 = 6.05). We say that 6.05 has more precision. If the value with the most decimal places is 2.23 and the lowest value is 1.5, a convenient starting point is 1.495 (1.5 – 0.005 = 1.495). If the value with the most decimal places is 3.234 and the lowest value is 1.0, a convenient starting point is 0.9995 (1.0 – 0.0005 = 0.9995). If all the data happen to be integers and the smallest value is two, then a convenient starting point is 1.5 (2 – 0.5 = 1.5). Also, when the starting point and other boundaries are carried to one additional decimal place, no data value will fall on a boundary.

2. Place frequency or relative frequency on the y-axis. Scale is important.

3. Draw bars as high as the frequency for each class interval within each boundary

The next two examples go into detail about how to construct a histogram using continuous data and how to create a histogram using discrete data.

Example 2.4

The following data are the heights (in inches to the nearest half inch) of 100 male semiprofessional soccer players. The heights are continuous data, since height is measured.

60; 60.5; 61; 61; 61.5; 63.5; 63.5; 63.5; 64; 64; 64; 64; 64; 64; 64; 64.5; 64.5; 64.5; 64.5; 64.5; 64.5; 64.5; 64.5; 66; 66; 66; 66; 66; 66; 66; 66; 66; 66; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 66.5; 67; 67; 67; 67; 67; 67; 67; 67; 67; 67; 67; 67; 67.5; 67.5; 67.5; 67.5; 67.5; 67.5; 67.5; 68; 68; 69; 69; 69; 69; 69; 69; 69; 69; 69; 69; 69.5; 69.5; 69.5; 69.5; 69.5; 70; 70; 70; 70; 70; 70; 70.5; 70.5; 70.5; 71; 71; 71; 72; 72; 72; 72.5; 72.5; 73; 73.5; 74

Create a relative frequency histogram using the frequency distribution that follows:

CLASSES	FREQUENCY (f)
60.0 – 61.9	5
62.0 – 63.9	3
64.0 – 65.9	15
66.0 – 67.9	40
68.0 – 69.9	17
70.0 – 71.9	12
72.0 – 73.9	7
74.0 – 75.9	1

Table 2.8 Frequency distribution of Heights

Solution 2.4

First find the relative frequencies and class boundaries for each class.

BOUNDARIES	RELATIVE FREQUENCY
59.95 – 61.95	.05
61.95 – 63.95	.03
63.95 – 65.95	.15
65.95 – 67.95	.4
67.95 – 69.95	.17
69.95– 71.95	.12
71.95 – 73.95	.07
73.95 – 75.95	.01

Second place the boundaries on the x-axis and the relative frequency on the y-axis.

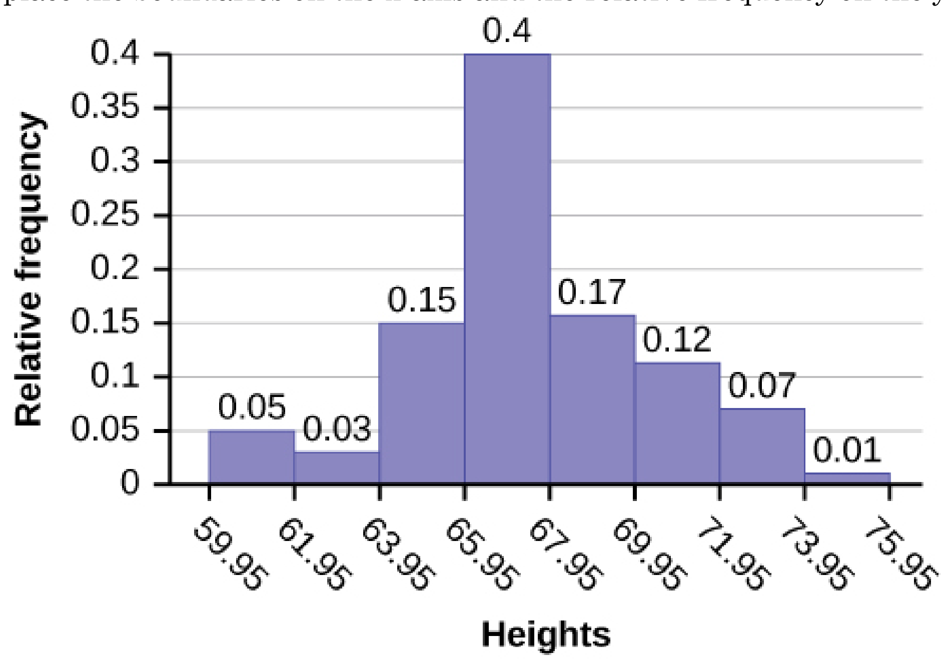


Figure 2.8

Example 2.5

The following data are the number of books bought by 50 part-time college students at the College of Lake County. The number of books is discrete data, since books are counted.

1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1
2; 2; 2; 2; 2; 2; 2; 2; 2; 2
3; 3; 3; 3; 3; 3; 3; 3; 3; 3; 3; 3; 3; 3; 3; 3
4; 4; 4; 4; 4; 4
5; 5; 5; 5; 5
6; 6;

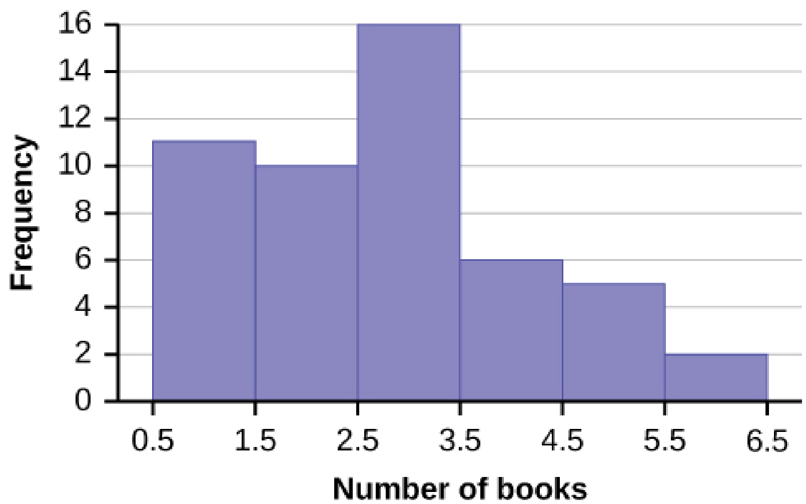
Construct a Histogram.

BOOKS	FREQUENCY (f)
1	11
2	10
3	16
4	6
5	5
6	2

Table 2.9 Ungrouped Frequency distribution of Books

Solution 2.5

BOUNDARIES	FREQUENCY
.5 – 1.5	11
1.5 – 2.5	10
2.5 – 3.5	16
3.5 – 4.5	6
4.5 – 5.5	5
5.5 – 6.5	2



Try It Σ

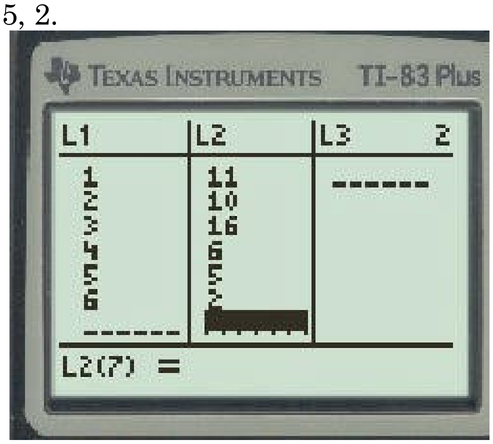
2.2 The following data are the shoe sizes of 50 male students. The sizes are continuous data since shoe size is measured. Construct a histogram and calculate the width of each bar or class interval. Suppose you choose six bars.

9; 9; 9.5; 9.5; 10; 10; 10; 10; 10; 10; 10.5; 10.5; 10.5; 10.5; 10.5; 10.5; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11; 11.5; 11.5; 11.5; 11.5; 11.5; 11.5; 11.5; 12; 12; 12; 12; 12; 12; 12; 12.5; 12.5; 12.5; 12.5; 14

Using the TI-83, 83+, 84, 84+ Calculator

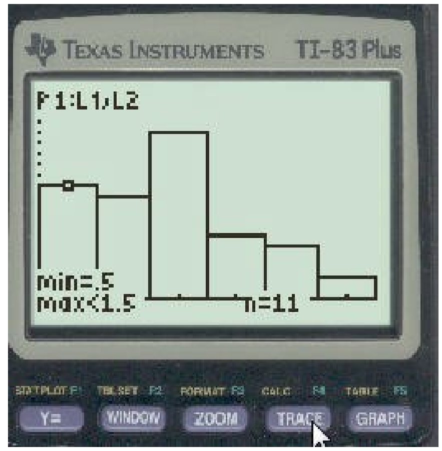
Go to **Appendix G**.
There are calculator instructions for entering data and for creating a customized histogram. Create the histogram for Example 2.5.

- Press **Y=**. Press CLEAR to delete any equations.
- Press **STAT** 1:EDIT. If L1 has data in it, arrow up into the name L1, press CLEAR and then arrow down. If necessary, do the same for L2.
- Into L1, enter 1, 2, 3, 4, 5, 6.
- Into L2, enter 11, 10, 16, 6, 5, 2.



- Press **WINDOW**. Set Xmin = .5, Xscl = (6.5 – .5)/6, Ymin = –1, Ymax = 20, Yscl = 1, Xres = 1.

- Press 2nd Y=. Start by pressing 4: Plotsoff ENTER.
- Press 2nd Y=. Press 1:Plot1. Press ENTER. Arrow down to TYPE. Arrow to the 3rd picture (histogram). Press ENTER.

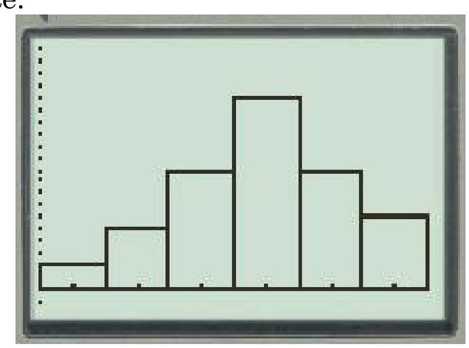


- Arrow down to Xlist: Enter L1 (2nd 1). Arrow down to Freq. Enter L2 (2nd 2). Press Graph. Use TRACE key and the arrow keys to examine the histogram.

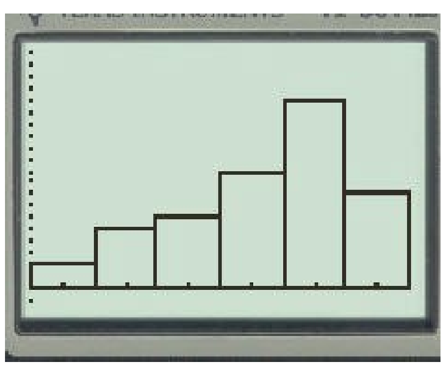
Descriptions of Histograms:

There are four descriptions of histograms. These descriptions, describe how the histogram is shaped.

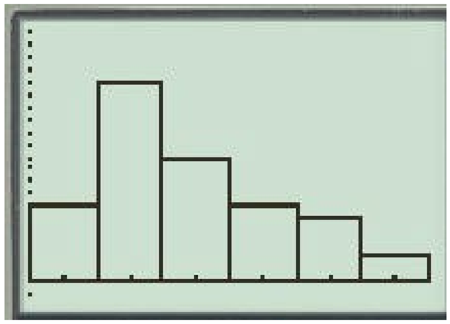
1. **Symmetrical:** the largest bar is in the center interval and the bars on each side of the center decrease about the same rate.



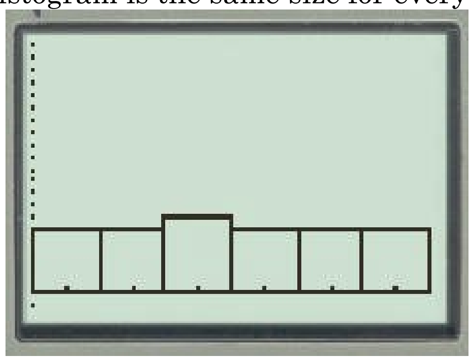
2. **Skewed Left**: the largest bar is closer to the right side and the bars start decreasing toward the left.



3. **Skewed Right:** the largest bar is closer to the left side and the bars start decreasing toward the right.



4. **Uniform**: each bar of the histogram is the same size for every interval



NOTE: Histograms will not always be exactly one of the four descriptions; therefore use adjectives to help describe the histogram. For example, "mostly symmetrical".

Stem-and-Leaf Graph:

One simple graph, the **stem-and-leaf** graph or stemplot, comes from the field of exploratory data analysis. It is a good choice when the data sets are small.

To create the **stem-and-leaf plot:**

1. Divide each observation of data into a stem and a leaf. The leaf consists of a final significant digit.

For example:

The number 23 has stem two and leaf three. The number 432 has stem 43 and leaf two. Likewise, the number 5,432 has stem 543 and leaf two. The decimal 9.3 has stem nine and leaf three.

2. Write the stems in a vertical line from smallest to largest. Draw a vertical line to the right of the stems. Then write the leaves in increasing order next to their corresponding stem.

Example 2.6

For Susan Dean's spring pre-calculus class, scores for the first exam were as follows:
33; 42; 49; 49; 53; 55; 55; 61; 63; 67; 68; 68; 69; 69; 72; 73; 74; 78; 80; 83; 88; 88; 88; 90; 92; 94; 94; 94; 94; 96; 100
Create a Stem and leaf plot.

Solution 2.6:

Stem	Leaf
3	3
4	2 9 9
5	3 5 5
6	1 3 7 8 8 9 9
7	2 3 4 8
8	0 3 8 8 8
9	0 2 4 4 4 4 6
10	0

Table 2.9 Stem-and- Leaf Graph

The stemplot shows that most scores fell in the 60s, 70s, 80s, and 90s. Eight out of the 31 scores or approximately 26% were in the 90s or 100, a fairly high number of A's.

Try It Σ

2.3 For the Park City basketball team, scores for the last 30 games were as follows (smallest to largest):

32; 32; 33; 34; 38; 40; 42; 42; 43; 44; 46; 47; 47; 48; 48; 48; 49; 50; 50; 51; 52; 52; 52; 53; 54; 56; 57; 57; 60; 61

Construct a stem plot for the data.

The stem and leaf plot is a quick way to graph data and gives an exact picture of the data. You want to look for an overall pattern and any outliers. An outlier is an observation of data that does not fit the rest of the data. It is sometimes called an extreme value. When you graph an outlier, it will appear not to fit the pattern of the graph. Some outliers are due to mistakes (for example, writing down 50 instead of 500) while others may indicate that something unusual is happening. It takes some background information to explain outliers, so we will cover them in more detail later.

Example 2.7

The data are the distances (in km) from a home to local supermarkets. Create a stemplot using the data:

1.1; 1.5; 2.3; 2.5; 2.7; 3.2; 3.3; 3.3; 3.5; 3.8; 4.0; 4.2; 4.5; 4.5; 4.7; 4.8; 5.5; 5.6; 6.5; 6.7; 12.3

Do the data seem to have any concentration of values? The leaves are to the right of the decimal.

Solution 2.7

The value 12.3 may be an outlier. Values appear to concentrate at three and four kilometers.

Stem	Leaf
1	1 5
2	3 5 7
3	2 3 3 5 8
4	0 2 5 5 7 8
5	5 6
6	5 7
7	
8	
9	
10	
11	
12	3

Table 2.10

Try It Σ

2.4 The following data show the distances (in miles) from the homes of off-campus statistics students to the college. Create a stem plot using the data and identify any outliers:

0.5; 0.7; 1.1; 1.2; 1.2; 1.3; 1.3; 1.5; 1.5; 1.7; 1.7; 1.8; 1.9; 2.0; 2.2; 2.5; 2.6; 2.8; 2.8; 2.8; 3.5; 3.8; 4.4; 4.8; 4.9; 5.2; 5.5; 5.7; 5.8; 8.0

Example 2.8

A side-by-side stem-and-leaf plot allows a comparison of the two data sets in two columns. In a side-by-side stem-and-leaf plot, two sets of leaves share the same stem. The leaves are to the left and the right of the stems. Table 2.11 and Table 2.12 show the ages of presidents at their inauguration and at their death. Construct a side- by-side stem-and-leaf plot using this data.

President	Age	President	Age	President	Age
Washington	57	A. Johnson	56	Truman	60
J. Adams	61	Grant	46	Eisenhower	62
Jefferson	57	Hayes	54	Kennedy	43
Madison	57	Garfield	49	L. Johnson	55
Monroe	58	Arthur	51	Nixon	56
J.Q. Adams	57	Cleveland	47	Ford	61
Jackson	61	B. Harrison	55	Carter	52
Van Buren	54	Cleveland	55	Reagan	69
W. H. Harrison	68	McKinley	54	G.H.W Bush	64
Tyler	51	T. Roosevelt	42	Clinton	47
Polk	49	Taft	51	G.W. Bush	54
Taylor	64	Wilson	56	Obama	47
Fillmore	50	Harding	55	Trump	70
Pierce	48	Coolidge	51	Biden	78
Buchanan	65	Hoover	54		
Lincoln	52	F. Roosevelt	51		

Table 2.11 Presidential Ages at Inauguration

President	Age	President	Age	President	Age
Washington	67	A. Johnson	66	Truman	88
J. Adams	90	Grant	63	Eisenhower	78
Jefferson	83	Hayes	70	Kennedy	46
Madison	85	Garfield	49	L. Johnson	64
Monroe	73	Arthur	56	Nixon	81
J.Q. Adams	80	Cleveland	71	Ford	93
Jackson	78	B. Harrison	67	Carter	
Van Buren	79	Cleveland	71	Reagan	93
W. H. Harrison	68	McKinley	58	G.H.W Bush	94
Tyler	71	T. Roosevelt	60	Clinton	
Polk	53	Taft	72	G.W. Bush	
Taylor	65	Wilson	67	Obama	
Fillmore	50	Harding	57	Trump	
Pierce	64	Coolidge	60	Biden	
Buchanan	77	Hoover	90		
Lincoln	56	F. Roosevelt	63		

Table 2.12 Presidential Ages at Death

Solution 2.8

Ages at Inauguration	Stem	Ages at Death
9 9 8 7 7 7 6 3 2	4	6 9
8 7 7 7 7 6 6 6 5 5 5 5 4 4 4 4 4 2 1 1 1 1 1 0	5	3 6 6 7 7 8
9 5 4 4 2 1 1 1 0	6	0 0 3 3 4 4 5 6 7 7 7 8
8 0	7	0 0 1 1 1 4 7 8 8 9
	8	0 1 3 5 8
	9	0 0 3 3 4

Line Graphs:

Another type of graph that is useful for specific data values is a line graph. In the particular **line graph** shown in Example 2.9 is for an ungrouped frequency distribution, the x-axis (horizontal axis) consists of data values and the y-axis (vertical axis) consists of frequency points. The frequency points are connected using line segments.

Example 2.9

In a survey, 40 mothers were asked how many times per week a teenager must be reminded to do his or her chores. The results are shown in Table 2.13 and in Figure 2.10.

# of times	FREQUENCY (f)
0	2
1	5
2	8
3	14
4	7
5	4

Table 2.13

Solution 2.9:

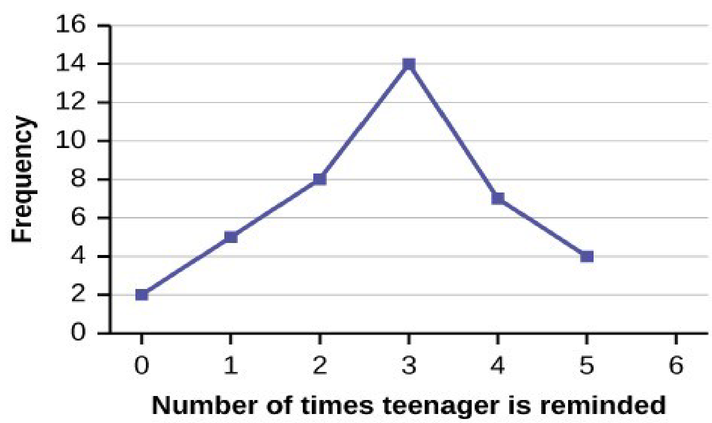


Figure 2.10

Try It Σ

2.5 In a survey, 40 people were asked how many times per year they had their car in the shop for repairs. The results are shown in Table 2.14. Construct a linegraph.

# of times	FREQUENCY (f)
0	7
1	10
2	14
3	9

Frequency Polygon:

Frequency polygons are analogous to line graphs, and just as line graphs make ***continuous*** data visually easy to interpret, so too do frequency polygons.

To construct **a frequency (relative frequency) polygon**:
1. Find the midpoints of each class and place them on the x-axis
2. Place frequency or relative frequency on the y-axis
3. Draw a line graph corresponding to each midpoint and frequency

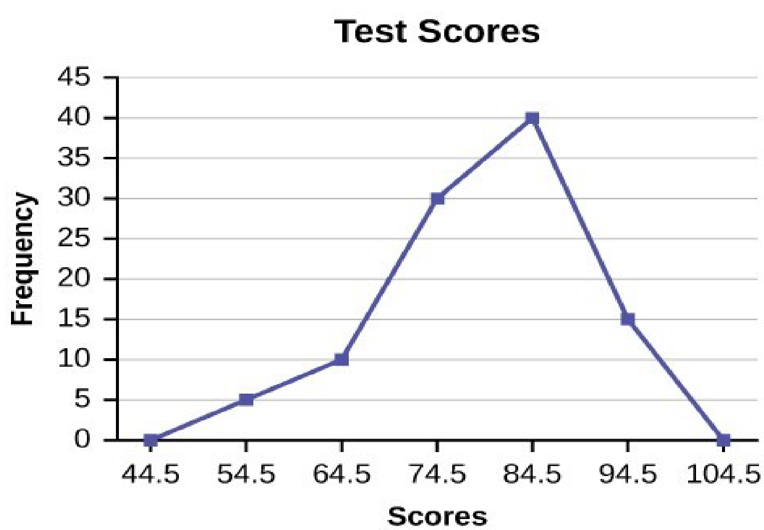
Example 2.10

Construct a frequency polygon using the distribution of Statistic Final Test Scores

CLASSES	FREQUENCY (f)
50 - 59	5
60 - 69	10
70 - 79	30
80 - 89	40
90 - 99	15

Solution 2.10:

First find the midpoints of each class. Midpoints: 54.5, 64.6, 74.5, 84.5, 94.5



The first label on the x-axis is 44.5. This represents an interval extending from 39.5 to 49.5. Since the lowest test score is 54.5, this interval is used only to allow the graph to touch the x-axis. The point labeled 54.5 represents the next interval or the first "real" interval from the table, and contains five scores. This reasoning is followed for each of the remaining intervals with the point 104.5 representing the interval from 99.5 to 109.5. Again, this interval contains no data and is only used so that the graph will touch the x-axis. Looking at the graph, we say that this distribution is skewed because one side of the graph does not mirror the other side.

Try It Σ

2.6 Construct a frequency polygon using the ages of President's inauguration in the following table.

Classes (Ages)	Frequency (f)
42 - 46	4
47 - 51	11
52 - 56	14
57 - 61	9
62 - 66	4
67 - 71	2

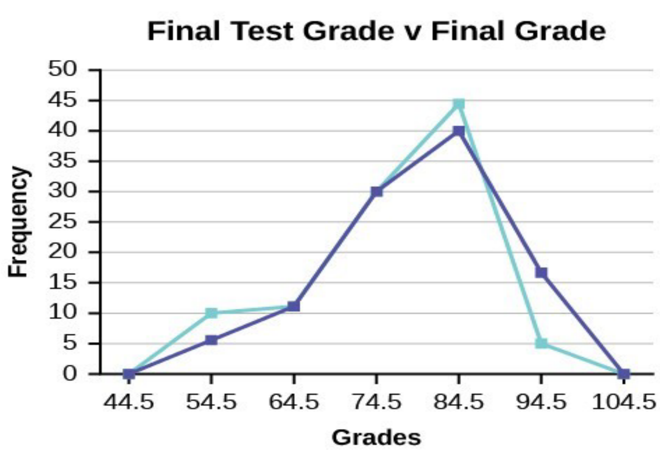
Frequency polygons are useful for comparing distributions. This is achieved by overlaying the frequency polygons drawn for different data sets.

Example 2.11

We will construct an overlay frequency polygon comparing the scores from Example 2.8 with the students' final numeric grade.

CLASSES	FREQUENCY (f)
50 - 59	10
60 - 69	10
70 - 79	30
80 - 89	45
90 - 99	5

Solution 2.11:



Suppose that we want to study the temperature range of a region for an entire month. Every day at noon we note the temperature and write this down in a log. A variety of statistical studies could be done with this data. We could find the mean or the median temperature for the month. We could construct a histogram displaying the number of days those temperatures reach a certain range of values. However, all of these methods ignore a portion of the data that we have collected.

One feature of the data that we may want to consider is that of time. Since each date is paired with the temperature reading for the day, we don't have to think of the data as being random. We can instead use the times given to impose a chronological order on the data. A graph that recognizes this ordering and displays the changing temperature as the month progresses is called a time series graph.

Constructing a Time Series Graph

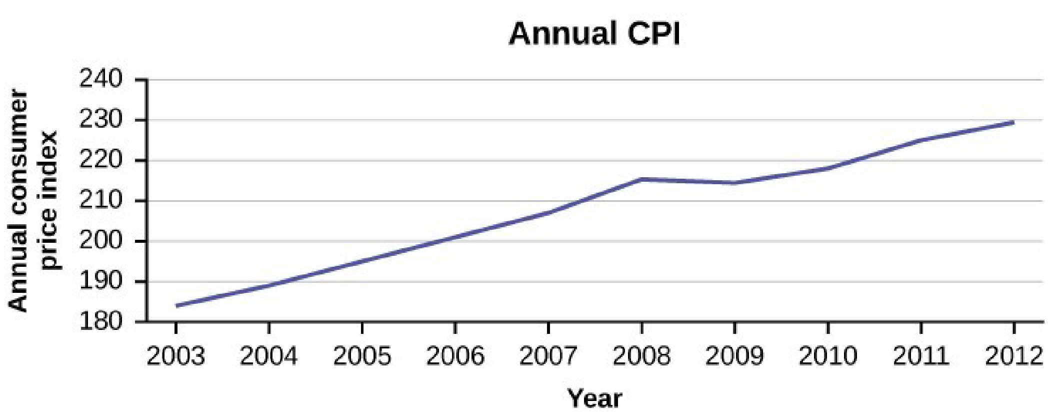
To construct a time series graph, we must look at both pieces of our paired data set. We start with a standard Cartesian coordinate system. The horizontal axis is used to plot the date or time increments, and the vertical axis is used to plot the values of the variable that we are measuring. By doing this, we make each point on the graph correspond to a date and a measured quantity. The points on the graph are typically connected by straight lines in the order in which they occur.

Example 2.12

The following data shows the Annual Consumer Price Index, each month, for ten years. Construct a time series graph for the Annual Consumer Price Index data only.

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Annual	184	188.9	195.3	201.6	207.342	215.303	214.537	218.056	224.939	229.594

Solution 2.12:



Try It Σ

2.7 The following table is a portion of a data set from www.worldbank.org. Use the table to construct a time series graph for CO2 emissions for the United States.

CO2 Emissions			
	Ukraine	**United Kingdom**	**United States**
2003	352,259	540,640	5,681,664
2004	343,121	540,409	5,790,761
2005	339,029	541,990	5,826,394
2006	327,797	542,045	5,737,615
2007	328,357	528,631	5,828,627
2008	323,657	522,247	5,656,839
2009	272,176	474,579	5,299,563

Uses of a Time Series Graph

Time series graphs are important tools in various applications of statistics. When recording values of the same variable over an extended period of time, sometimes it is difficult to discern any trend or pattern. However, once the same data points are displayed graphically, some features jump out. Time series graphs make trends easy to spot.

KEY TERMS

Bar Graphs a representation of categorical data where horizontal axis contains the qualitative categories, and the vertical axis represents the frequency of each category. The height of the bars represents the amount of data in the category.

Frequency Polygon looks like a line graph but uses intervals to display ranges of large amounts of data.

Frequency Table a data representation in which grouped data is displayed along with the corresponding frequencies.

Frequency the number of times a value of the data occurs.

Histogram a graphical representation in of the distribution of data in a data set; the horizontal axis represents the data, and the vertical axis represents the frequencies or relative frequencies. The graph consists of contiguous rectangles.

Interval also called a class interval; an interval represents a range of data and is used when displaying large data sets.

Paired Data Set two data sets that have a one-to-one relationship so that:
- both data sets are the same size, and
- each data point in one data set is matched with exactly one point from the other set.

Pie Graphs a circular graph that shows how large each category is in relation to the whole. Created from the frequency distribution and using relative frequencies, the central angles are calculated by multiplying the relative frequency by 360 degrees.

Qualitative Data categorical data consisting of labels or descriptions.

Quantitative Continuous Data measurable data that is numerical and can take on any value in a given range of numbers.

Quantitative Discrete Data measurable data that is numerical that can only take on particular values and cannot take on the values in between.

Relative Frequency the ratio of the number of times a value of the data occurs in the set of all outcomes to the number of all outcomes.

Time Series the same variable is measured over a period and plotted on the x-y coordinate plane where time is on the x-axis and the variable in the y-axis.

EXERCISES for CHAPTER 2

For exercises #1 – 4, use the following frequency distribution

Number of times in store	**Frequency**
1	4
2	10
3	16
4	6
5	4

1. Find the boundaries for the second class.

2. Find the relative frequency for the third class.

3. What is the sample size?

4. What is the cumulative frequency for the fourth class?

5. The following data represent the number of potholes on 35 randomly selected1-mile stretches of highway around a particular city. Construct an ungrouped frequency distribution of the data.

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

For exercises #6 – 11, use the following frequency distribution

Temperature (°F)	**Frequency**
32 - 36	1
37 - 41	3
42 - 46	5
47 - 51	11
52 - 56	7
57 – 61	7
62 - 66	1

6. Find the class width.

7. Find the class midpoints.

8. Find the class boundaries.

9. Find the relative frequencies.

10. Find the cumulative frequencies.

11. Find the sample size.

For exercises #12 – 17, use the following frequency distribution

Time (minutes)	Frequency
70 - 78	14
79 - 87	15
88 - 96	16
97 - 105	18
106 - 114	22
115 - 123	7
124 - 132	3

12. Find the class width.

13. Find the class midpoints.

14. Find the class boundaries.

15. Find the relative frequencies.

16. Find the cumulative frequencies.

17. Find the sample size.

18. Use the given minimum and maximum data entries, and the number of classes, to find the class width, the lower class limits, and the upper class limits.
a. minimum = 12, maximum = 68, 6 classes
b. minimum = 28, maximum = 75, 6 classes
c. minimum = 5, maximum = 92, 8 classes
d. minimum = 30, maximum = 83, 8 classes

19. The following data represents the number of miles a full tank of gas allows for a sample of vehicles inspected. Create a grouped frequency distribution with 7 classes.

260	271	236	244	279	296	284	299	288
288	247	256	338	360	341	333	261	266
287	296	313	311	307	307	299	303	277
283	304	305	288	290	288	289	297	299
332	330	309	328	307	328	285	291	295
298	306	315	310	318	318	320	333	321
323	324	327						

20. The data represent the time, in minutes, spent reading a political blog in a day. Construct a frequency distribution using 5 classes. In the table, include the midpoints, relative frequencies, and cumulative frequencies. Which class has the greatest frequency and which has the least frequency?

34 18 10 44 41 47 0 33 49 45
16 10 27 24 44 46 21 25 45 23

21. Construct a frequency distribution for the given data set using 6 classes. In the table, include the midpoints, relative frequencies, and cumulative frequencies. Which class has the greatest frequency and which has the least frequency?

Amount (in dollars) spent on books for a semester
508 488 430 49 278 542 451 366 451 531 196 115
118 362 434 399 148 30 287 236 185 210 178 105
533 236 477 329

22. The students in Ms. Ramirez's math class have birthdays in each of the four seasons. The table below shows the four seasons, the number of students who have birthdays in each season, and the percentage (%) of students in each group. Construct a bar graph showing the number of students.

Seasons	**Number of Students**	**Percentage**
Spring	8	24%
Summer	9	26%
Autumn	11	32%
Winter	6	18%

23. Using the data from Mrs. Ramirez's math class supplied in Exercise 22, construct a pie graph showing the percentages.

24. David County has six high schools. Each school sent students to participate in a county-wide science competition. The table below shows the percentage breakdown of competitors from each school, and the percentage of the entire student population of the county that goes to each school. Construct a bar graph and pie graph that shows the population percentage of competitors from each school.

High School	**Science Competition Percentage**	**Overall Student Population Percentage**
Alabaster	28.9%	8.6%
Concordia	7.6%	23.2%
Genoa	12.1%	15.0%
Mocksville	18.5%	14.3%
Tynneson	24.2%	10.1%
West End	8.7%	28.8%

25. Use the data from the David County science competition supplied in 24. Construct a bar graph that shows population percentage of students at each school.

For exercises #26 - 29, create a stem/leaf plot for the given data.

26. The miles per gallon rating for 30 cars are shown below (lowest to highest).

19, 19, 19, 20, 21, 21, 25, 25, 25, 26, 26, 28, 29, 31, 31,
32, 32, 33, 34, 35, 36, 37, 37, 38, 38, 38, 38, 41, 43, 43

27. The height in feet of 25 trees is shown below (lowest to highest).

25, 27, 33, 34, 34, 34, 35, 37, 37, 38, 39, 39, 39,
40, 41, 45, 46, 47, 49, 50, 50, 53, 53, 54, 54

28. The following data are the prices of different laptops at an electronics store. Round each value to the nearest multiple of ten.

249, 249, 260, 265, 265, 280, 299, 299, 309, 319, 325, 326, 350,
350, 350, 365, 369, 389, 409, 459, 489, 559, 569, 570, 610

29. The following data are the daily high temperatures in a town for one month.

61, 61, 62, 64, 66, 67, 67, 67, 68, 69, 70, 70, 70, 71, 71,
72, 74, 74, 74, 75, 75, 75, 76, 76, 77, 78, 78, 79, 79, 95

30. In a survey, 40 people were asked how many times they visited a store before making a major purchase. Construct a frequency polygon. The results are shown in the table:

Number of times in store	**Frequency**
1	4
2	10
3	16
4	6
5	4

31. In a survey, several people were asked how many years it has been since they purchased a mattress. Construct a histogram. The results are shown below.

Years Since Last Purchase	**Frequency**
0	2
1	8
2	13
3	22
4	16
5	9

32. Several children were asked how many TV shows they watch each day. Construct a relative frequency polygon. The results of the survey are shown in the table below:

Number of TV Shows	Frequency
0	12
1	18
2	36
3	7
4	2

33. Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars. Complete the table.

Data Values (# cars)	Frequency	Relative Frequency	Cumulative Relative Frequency

34. Refer to the table for problem 33. Construct a histogram for the given data.

35. Construct a frequency polygon and histogram for each of the following frequency tables:

a.

Pulse Rates for Women	**Frequency**
60 - 69	12
70 - 79	14
80 – 89	11
90 – 99	1
100 – 109	1
110 – 119	0
120 – 129	1

b.

Actual Speed in a 30 mph Zone	**Frequency**
42 – 45	25
46 – 49	14
50 – 53	7
54 – 57	3
58 – 61	1

c.

Tar (in mg) of Nonfiltered Cigarettes	**Frequency**
10 – 13	1
14 – 17	0
18 – 21	15
22 – 25	7
26 – 29	2

36. Construct a relative frequency polygon and relative frequency histogram from the frequency distribution for the 50 highest ranked countries for depth of hunger.

Depth of Hunger	**Frequency**
230 - 259	21
260 - 289	13
290 - 319	6
320 - 349	7
350 - 379	1
380 - 409	1
410 - 439	1

37. Use the two frequency tables to compare the life expectancy of men and women from 20 randomly selected countries. Include an overlaid frequency polygon and discuss the shapes of the distributions, the center, the spread, and any outliers. What can we conclude about the life expectancy of women compared to men?

Life Expectancy at Birth – Women	Frequency
49 – 55	3
56 – 62	3
63 – 69	1
70 – 76	3
77 – 83	8
84 - 90	2
Life Expectancy at Birth – Men	Frequency
49 – 55	3
56 – 62	3
63 – 69	1
70 – 76	1
77 – 83	7
84 - 90	5

38. Construct a times series graph for (a) the number of male births, (b) the number of female births, and (c) the total number of births.

Sex/Yr	**1855**	**1856**	**1857**	**1858**	**1859**	**1860**	**1861**
Female	49,545	49,582	50,257	50,234	51,915	51,220	52,403
Male	47,804	52,239	53,158	53,694	54,628	54,409	54,606
Total	93,349	101,821	103,415	104,018	106,543	105,629	107,009

39. The following data sets list full time police per 100,000 citizens along with homicides per 100,000 citizens for the city of Detroit, Michigan during the period from 1961 to 1972.

Year	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
Police	260.4	269.8	272	273	272.5	261.3	268.9	296	319.9	341.4	356.6	376.7
Homicides	8.6	8.9	8.5	8.9	13.1	14.6	21.4	28	31.5	37.4	46.3	47.2

a. Construct a double time series graph using a common *x*-axis for both sets of data.
b. Which variable increased the fastest? Explain.
c. Did Detroit's increase in police officers have an impact on the murder rate? Explain.

REFERENCES

2.2 Graphs for Qualitative Data

Burbary, Ken. *Facebook Demographics Revisited – 2001 Statistics,* 2011. Available online at http://www.kenburbary.com/2011/03/facebook-demographics-revisited-2011-statistics-2/ (accessed August 21, 2013).

"9th Annual AP Report to the Nation." CollegeBoard, 2013. Available online at http://apreport.collegeboard.org/goals-and-findings/promoting-equity (accessed September 13, 2013).

"Overweight and Obesity: Adult Obesity Facts." Centers for Disease Control and Prevention. Available online at http://www.cdc.gov/obesity/data/adult.html (accessed September 13, 2013).

2.3 Graphs for Quantitative Data

Data on annual homicides in Detroit, 1961–73, from Gunst & Mason's book 'Regression Analysis and its Application', Marcel Dekker

"Timeline: Guide to the U.S. Presidents: Information on every president's birthplace, political party, term of office, and more." Scholastic, 2013. Available online at http://www.scholastic.com/teachers/article/timeline-guide-us-presidents (accessed April 3, 2013).

"Presidents." Fact Monster. Pearson Education, 2007. Available online at http://www.factmonster.com/ipka/A0194030.html (accessed April 3, 2013).

"Food Security Statistics." Food and Agriculture Organization of the United Nations. Available online at http://www.fao.org/economic/ess/ess-fs/en/ (accessed April 3, 2013).

"Consumer Price Index." United States Department of Labor: Bureau of Labor Statistics. Available online at http://data.bls.gov/pdq/SurveyOutputServlet (accessed April 3, 2013).

"CO2 emissions (kt)." The World Bank, 2013. Available online at http://databank.worldbank.org/data/home.aspx (accessed April 3, 2013).

"Births Time Series Data." General Register Office For Scotland, 2013. Available online at http://www.gro-scotland.gov.uk/statistics/theme/vital-events/births/time-series.html (accessed April 3, 2013).

"Demographics: Children under the age of 5 years underweight." Indexmundi. Available online at http://www.indexmundi.com/g/r.aspx?t=50&v=2224&aml=en (accessed April 3, 2013).

Gunst, Richard, Robert Mason. *Regression Analysis and Its Application: A Data-Oriented Approach*. CRC Press: 1980.

"Overweight and Obesity: Adult Obesity Facts." Centers for Disease Control and Prevention. Available online at http://www.cdc.gov/obesity/data/adult.html (accessed September 13, 2013).

CHAPTER 2 SOLUTIONS:

1) 1.5 – 2.5

3) $n = \Sigma f = 40$

5) frequency distribution

7) midpoints: 34, 39, 44, 49, 54, 59, 64

# of potholes	**f**
1	12
2	4
3	5
4	5
5	3
6	3
7	3

9) r.f: 2.9%, 8.6%, 14.3%, 31.4%, 20%, 20%, 2.9%

11) $n = \Sigma f = 35$

13) midpoints: 74, 83, 92, 101, 110, 119, 128

15) r.f: .147, .158, .168, .189, .232, .074, .032

17) $n = \Sigma f = 95$

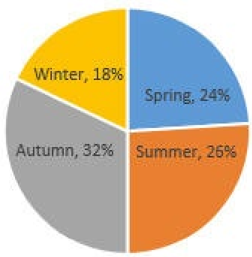
19) grouped frequency distribution with class width = 18

Classes	Frequency
236 – 253	3
254 – 271	5
272 – 289	11
290 – 307	17
308 – 325	11
326 – 343	9
344 – 361	1

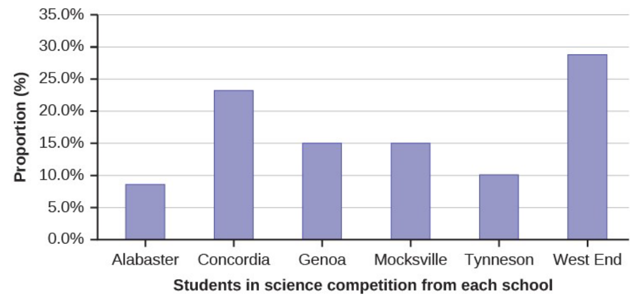
21) grouped frequency distribution with class width = 86

Classes	Frequency	Midpoint	Rel. Freq.	Cum. Freq.
30 – 115	4	72.5	14.3%	4
116 – 201	5	158.5	17.9%	9
202 – 287	5	244.5	17.9%	14
288 – 373	3	330.5	10.7%	17
374 – 459	5	416.5	17.9%	22
460 – 545	6	502.5	21.4%	28

23)



25)



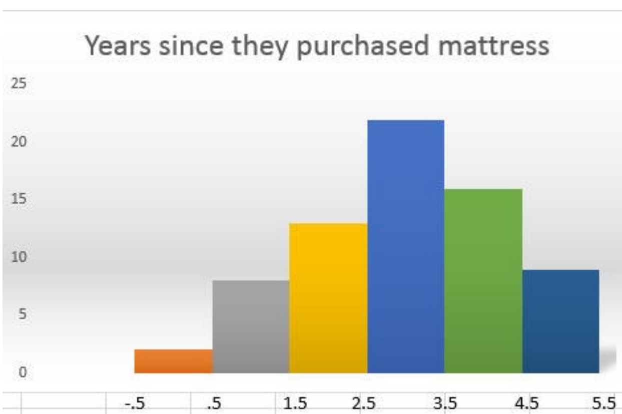
27)

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

29)

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

31)

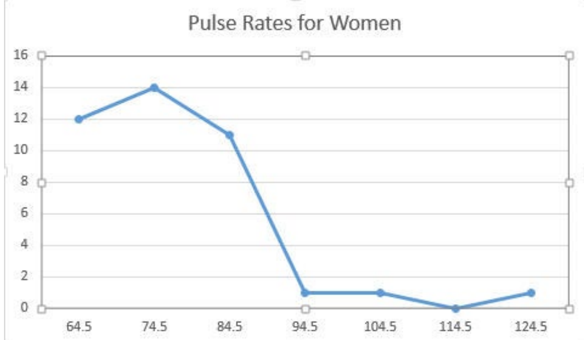


33)

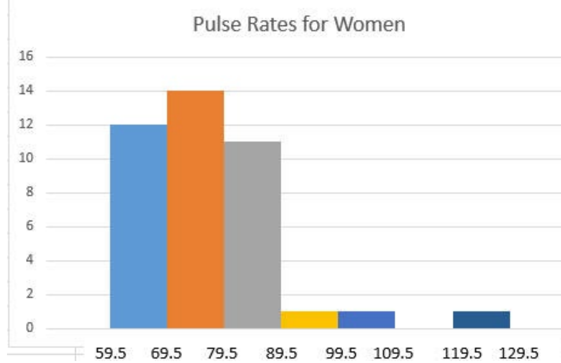
Cars	Freq.	Rel Freq	Cum. Freq
3	14	21.5%	14
4	19	29.2%	33
5	12	18.5%	45
6	9	13.8%	54
7	11	16.9%	65

35)

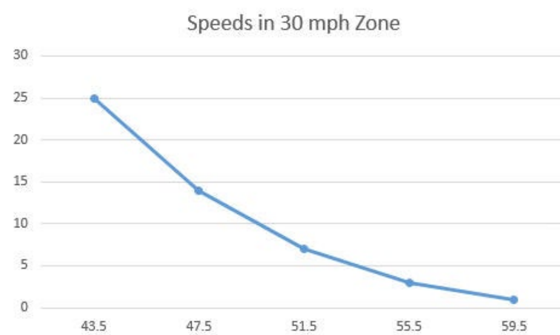
a. Frequency Polygon

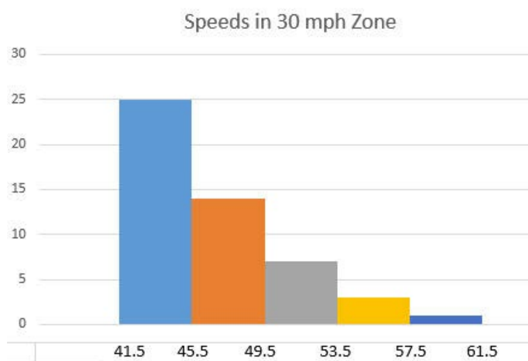


Histogram

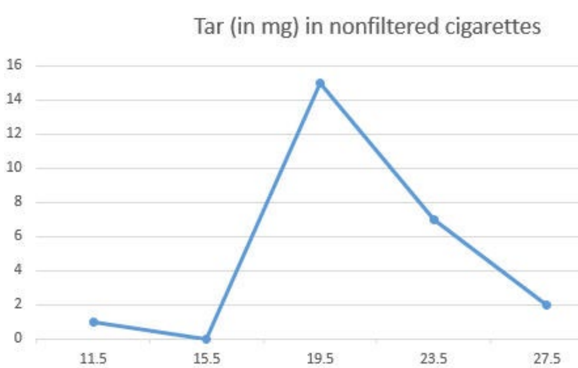


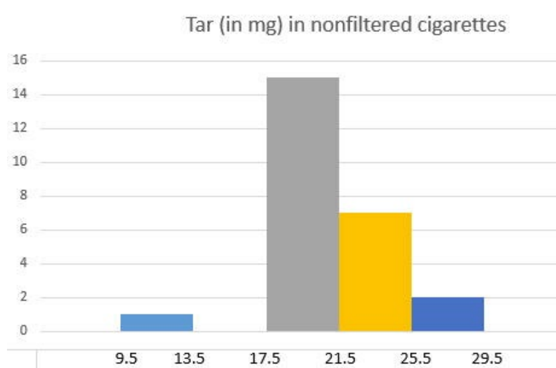
b.



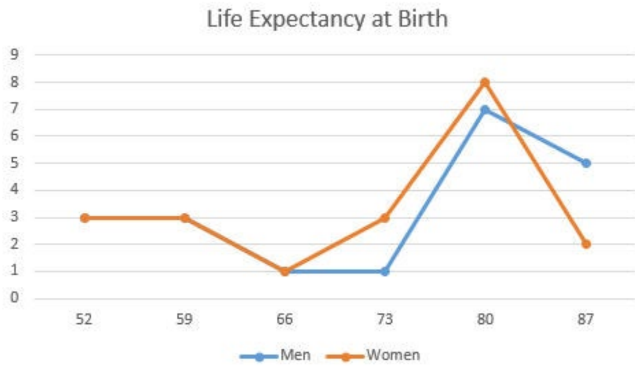


c.



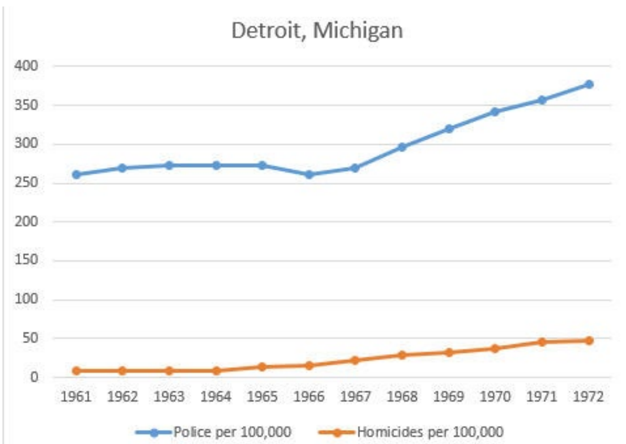


37)



Both skewed left. Similar life expectancy pattern for both men and women

39) Police variable increased faster. The number of police did not impact the number of homicides. Homicides continued to increase.



This page is purposely left blank.