Elementary Statistics Using the TI 83 84 Plus Calculator 5th edition Triola Solutions Manual

SKU: 9780134686943-SOLUTIONS

Chapter 2: Exploring Data with Tables and Graphs

Section 2-1: Frequency Distributions for Organizing and Summarizing Data

1. The table summarizes 50 service times. It is not possible to identify the exact values of all of the original times.

2. The classes of 60–120, 120–180, …, 300–360 overlap, so it is not always clear which class we should put a value in. For example, the value of 120 could go in the first class or the second class. The classes should be mutually exclusive.

3.

Time (sec)	Relative Frequency
60–119	14%
120–179	44%
180–239	28%
240–299	4%
300–359	10%

4. The sum of the relative frequencies is 125%, but it should be 100%, with a small round off error. All of the relative frequencies appear to be roughly the same, but if they are from a normal distribution, they should start low, reach a maximum, and then decrease.

5. Class width: 10
Class midpoints: 24.5, 34.5, 44.5, 54.5, 64.5, 74.5, 84.5
Class boundaries: 19.5, 29.5, 39.5, 49.5, 59.5, 69.5, 79.5, 89.5
Number: 87

6. Class width: 10
Class midpoints: 24.5, 34.5, 44.5, 54.5, 64.5, 74.5
Class boundaries: 19.5, 29.5, 39.5, 49.5, 59.5, 69.5, 79.5
Number: 87

7. Class width: 100
Class midpoints: 49.5, 149.5, 249.5, 349.5, 449.5, 549.5, 649.5
Class boundaries: –0.5, 99.5, 199.5, 299.5, 399.5, 499.5, 599.5, 699.5
Number: 153

8. Class width: 100
Class midpoints: 149.5, 249.5, 349.5, 449.5, 549.5
Class boundaries: 99.5, 199.5, 299.5, 399.5, 499.5, 599.5
Number: 147

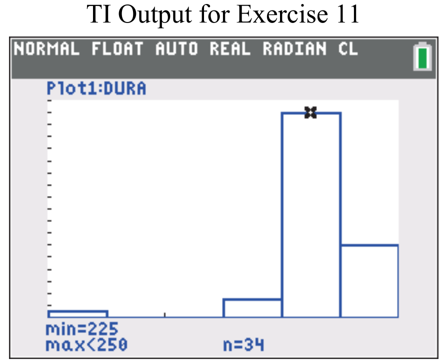
9. No. The maximum frequency is in the second class instead of being near the middle, so the frequencies below the maximum do not mirror those above the maximum.

10. Yes. The frequencies start low, reach a maximum of 36, and then decrease. The values below the maximum are very roughly a mirror image of those above it.

11.

Duration (sec)	Frequency
125–149	1
150–174	0
175–199	0
200–224	3
225–249	34
250–274	12

TI Output for Exercise 11

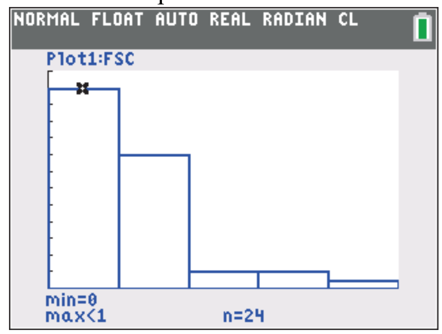


[125, 275] by [0, 36]

12. The intensities do not appear to have a normal distribution.

Tornado F-Scale	Frequency
0	24
1	16
2	2
3	2
4	1

TI Output for Exercise 12

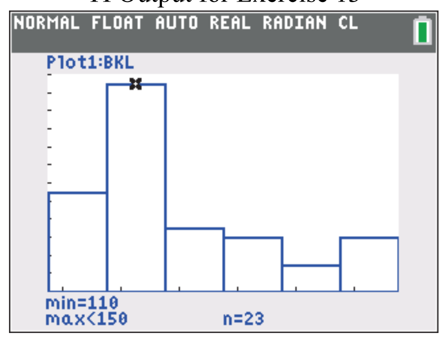


[0, 5] by [0, 26]

13.

Burger King Lunch Service Times (sec)	Frequency
70–109	11
110–149	23
150–189	7
190–229	6
230–269	3
270–309	6

TI Output for Exercise 13

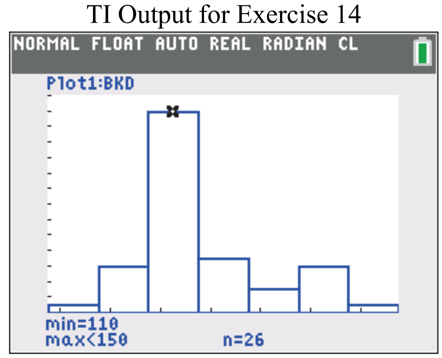


[70,310] by [0, 24]

14.

Burger King Dinner Service Times (sec)	Frequency
30–69	1
70–109	6
110–149	26
150–189	7
190–229	3
230–269	6
270–309	1

TI Output for Exercise 14

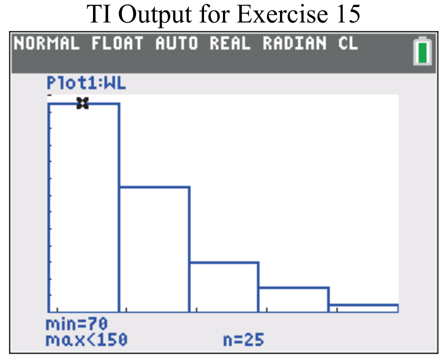


[30, 310] by [0, 28]

15. The distribution does not appear to be a normal distribution.

Wendy's Lunch Service Times (sec)	Frequency
70–149	25
150–229	15
230–309	6
310–389	3
390–469	1

TI Output for Exercise 15

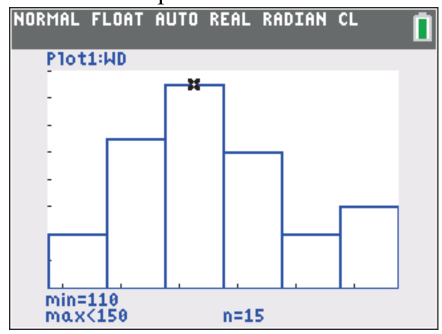


[70, 470] by [0, 26]

16. The distribution does appear to be a normal distribution.

Wendy's Dinner Service Times (sec)	Frequency
30–69	4
70–109	11
110–149	15
150–189	10
190–229	4
230–269	6

TI Output for Exercise 16

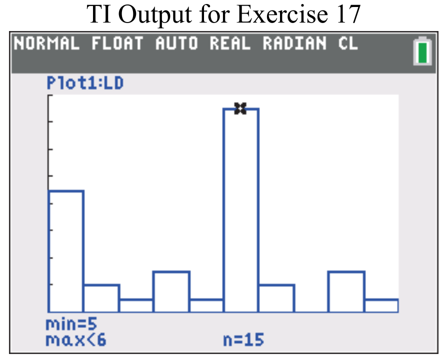


[30, 270] by [0, 16]

17. Because there are disproportionately more 0s and 5s, it appears that the heights were reported instead of measured. Consequently, it is likely that the results are not very accurate.

Last Digit	Frequency
0	9
1	2
2	1
3	3
4	1
5	15
6	2
7	0
8	3
9	1

TI Output for Exercise 17

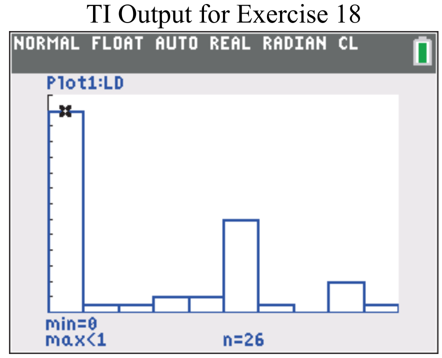


[0, 10] by [0, 16]

18. Because there are disproportionately more 0s and 5s, it appears that the weights were reported instead of measured. Consequently, it is likely that the results are not very accurate.

Last Digit	Frequency
0	26
1	1
2	1
3	2
4	2
5	12
6	1
7	0
8	4
9	1

TI Output for Exercise 18



[0, 10] by [0, 28]

19. The actresses appear to be younger than the actors.

Age When Oscar Was Won	Relative Frequency (Actresses)	Relative Frequency (Actors)
20 – 29	33.3%	1.1%
30 – 39	39.1%	32.2%
40 – 49	16.1%	41.4%
50 – 59	3.4%	17.2%
60 – 69	5.7%	6.9%
70 – 79	1.1%	1.1%
80 – 89	1.1%	

20. There do appear to be differences, but overall they are not very substantial differences.

Blood Platelet Count	Males	Females
0–99	0.7%	
100–199	33.3%	17.0%
200–299	58.8%	62.6%
300–399	6.5%	19.0%
400–499	0%	0%
500–599	0%	1.4%
600–699	0.7%	

21.

Age (years) of Best Actress When Oscar Was Won	Cumulative Frequency
Less than 30	29
Less than 40	63
Less than 50	77
Less than 60	80
Less than 70	85
Less than 80	86
Less than 90	87

22.

Age (years) of Best Actor When Oscar Was Won	Cumulative Frequency
Less than 30	1
Less than 40	29
Less than 50	65
Less than 60	80
Less than 70	86
Less than 80	87

23. No. The highest relative frequency of 24.8% is not much higher than the others.

Adverse Reaction	Relative Frequency
Headache	23.6%
Hypertension	8.7%
Upper Resp. Tract Infection	24.8%
Nasopharyngitis	21.1%
Diarrhea	21.9%

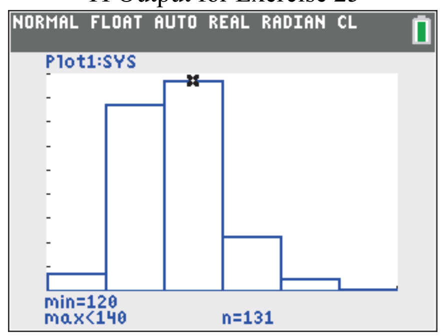
24. Yes, it appears that births occur on the days of the week with frequencies that are about the same.

Day	Relative Frequency
Monday	13.0%
Tuesday	16.5%
Wednesday	18.0%
Thursday	14.3%
Friday	14.3%
Saturday	10.8%
Sunday	13.3%

25. Yes, the frequency distribution appears to be a normal distribution.

Systolic Blood Pressure (mm Hg)	Frequency
80–99	11
100–119	116
120–139	131
140–159	34
160–179	7
180–199	1

TI Output for Exercise 25

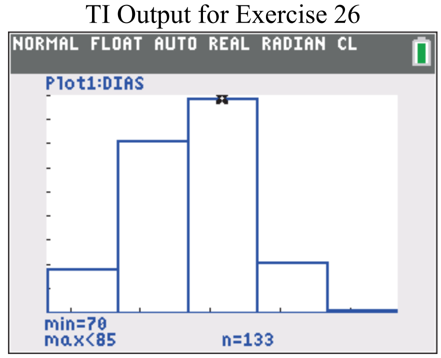


[80, 200] by [0, 135]

26. Yes, the frequency distribution appears to be a normal distribution.

Diastolic Blood Pressure (mm Hg)	Frequency
40–54	27
55–69	107
70–84	133
85–99	31
100–114	2

TI Output for Exercise 26

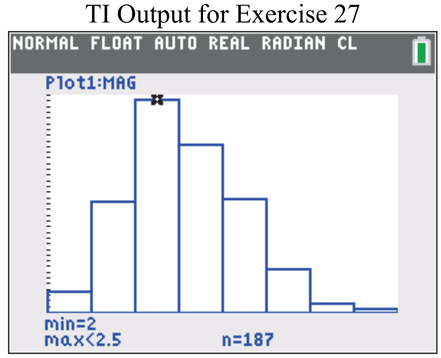


[40, 115] by [0, 135]

27. Yes, the frequency distribution appears to be a normal distribution.

Magnitude	Frequency
1.00–1.49	19
1.50–1.99	97
2.00–2.49	187
2.50–2.99	147
3.00–3.49	100
3.50–3.99	38
4.00–4.49	8
4.50–4.99	4

TI Output for Exercise 27

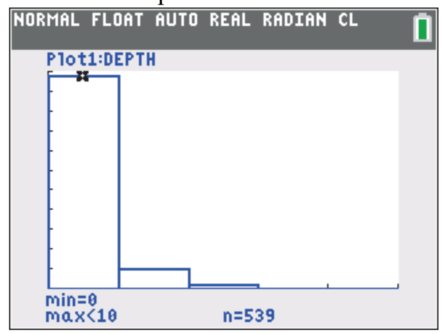


[0, 5] by [0, 190]

28. No, the frequency distribution does not appear to be a normal distribution.

Depth (km)	Frequency
0.0–9.9	539
10.0–19.9	49
20.0–29.9	10
30.0–39.9	1
40.0–49.9	1

TI Output for Exercise 28



[0, 50] by [0, 550]

29. An outlier can dramatically increase the number of classes.

Weight (lb)	With Outlier	Without Outlier
200–219	6	6
220–239	5	5
240–259	12	12
260–279	36	36
280–299	87	87
300–319	28	28
320–339	0	
340–359	0	
360–379	0	
380–399	0	
400–419	0	
420–439	0	
440–459	0	
460–479	0	
480–499	0	
500–519	1	

Section 2-2: Histograms

1. The histogram should be bell-shaped.

2. Not necessarily. Because the sample subjects themselves chose to be included, the voluntary response sample might not be representative of the population.

3. With a data set that is so small, the true nature of the distribution cannot be seen with a histogram.

4. The outlier will result in a single bar that is far away from all of the other bars in the histogram, and the height of that bar will correspond to a frequency of 1.

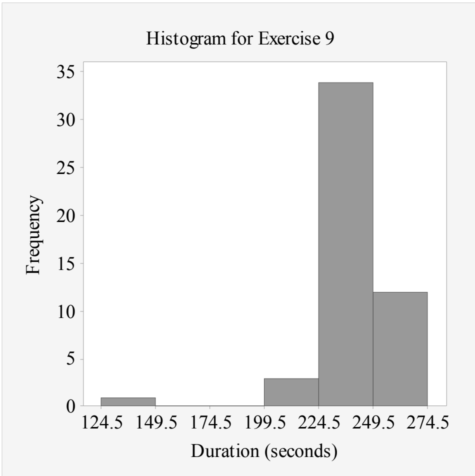
5. 40

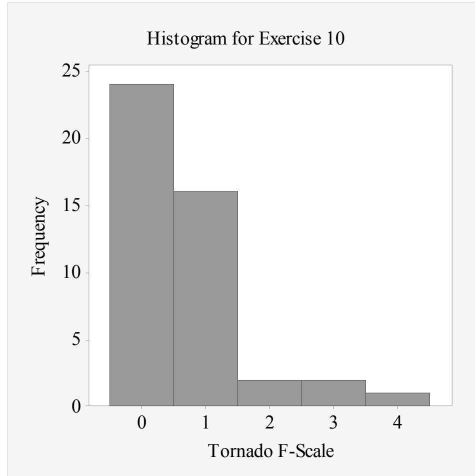
6. Approximate values: Class width: 0.1 gram, lower limit of first class: 5.5 grams, upper limit of first class: 5.6 grams

7. The shape of the graph would not change. The vertical scale would be different, but the relative heights of the bars would be the same.

8. 40 of the quarters are “pre-1964” made with 90% silver and 10% copper, and the other 40 quarters are “post-1964” made with a copper-nickel alloy. The histogram depicts weights from two different populations of quarters.

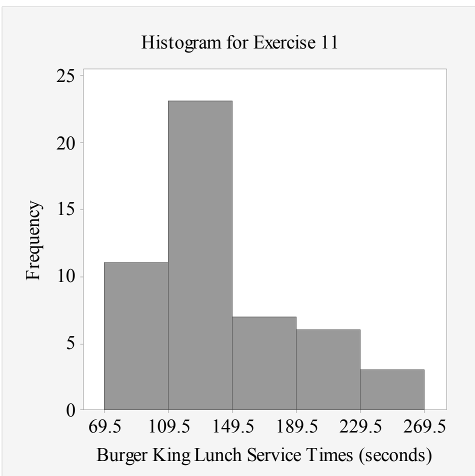
9. Because it is far from being bell-shaped, the histogram does not appear to depict data from a population with a normal distribution.

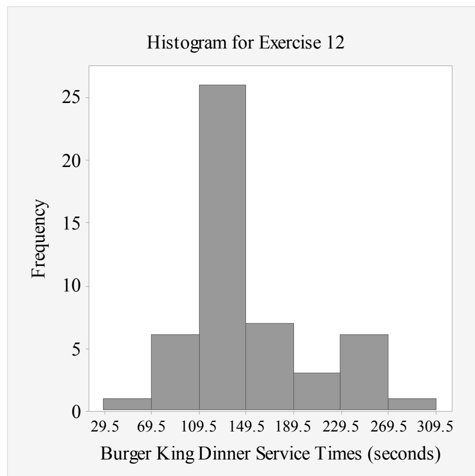




10. The histogram appears to be skewed to the right (or positively skewed).

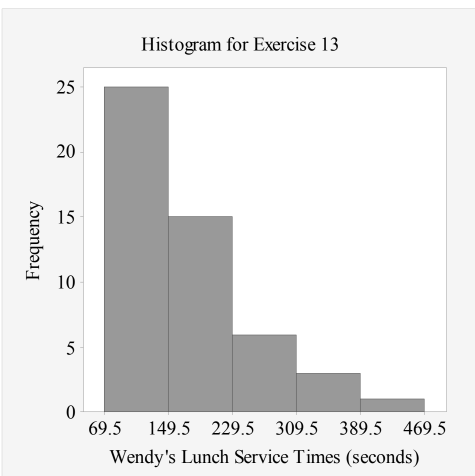
11. The histogram appears to be skewed to the right (or positively skewed).

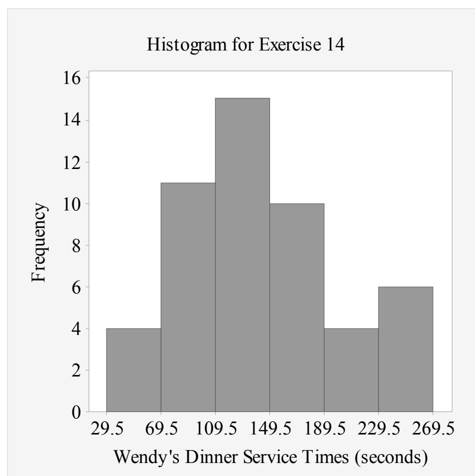




12. Because the histogram isn't close enough to being bell-shaped, it does not appear to depict data from a population with a normal distribution.

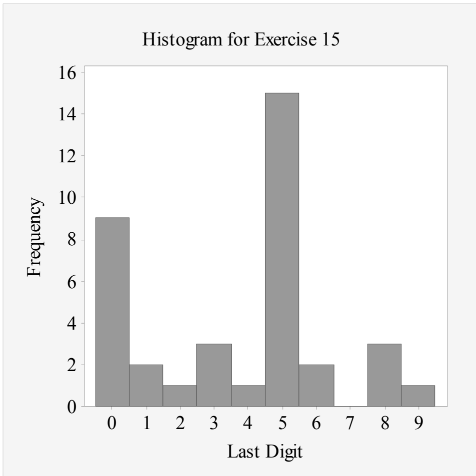
13. The histogram appears to be skewed to the right (or positively skewed).

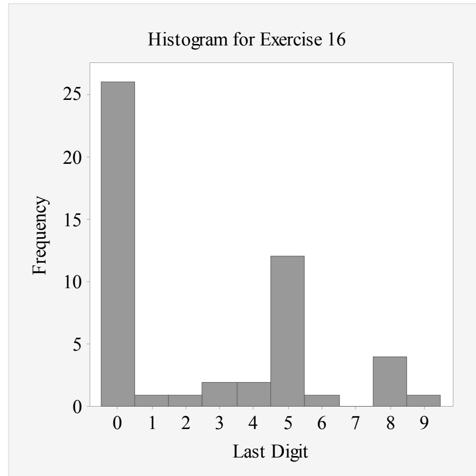




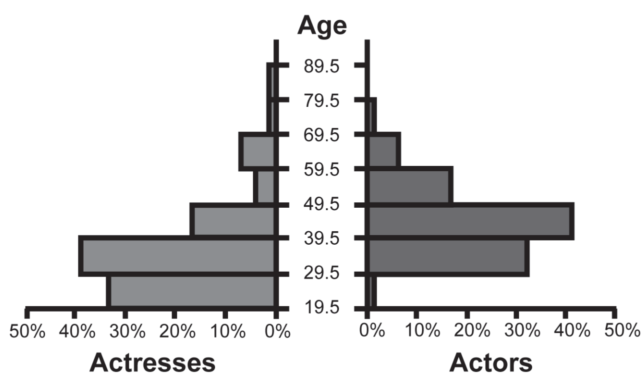
14. Because the histogram is not close enough to being bell-shaped, it does not appear to depict data from a population with a normal distribution.

15. The digits 0 and 5 appear to occur more often than the other digits, so it appears that the heights were reported and not actually measured. This suggests that the data might not be very useful.





16. The digits 0 and 5 appear to occur more often than the other digits, so it appears that the weights were reported and not actually measured. This suggests that the data might not be very useful.
17. The ages of actresses are lower than the ages of actors.



18. Only part (c) appears to represent data from a normal distribution. Part (a) has a systematic pattern that is not that of a straight line, part (b) has points that are not close to a straight-line pattern, and part (d) is really bad because it shows a systematic pattern and points that are not close to a straight-line pattern.

Section 2-3: Graphs That Enlighten and Graphs That Deceive

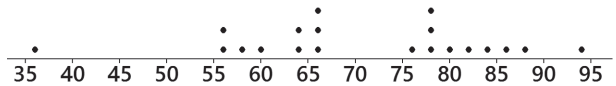
1. The data set is too small for a graph to reveal important characteristics of the data. With such a small data set, it would be better to simply list the data or place them in a table.

2. No. If the sample is a bad sample, such as one obtained from voluntary responses, there are no graphs or other techniques that can be used to salvage the data.

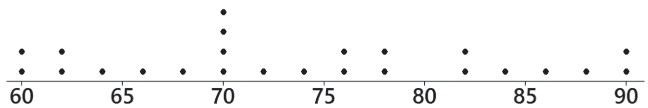
3. No. Graphs should be constructed in a way that is fair and objective. The readers should be allowed to make their own judgments, instead of being manipulated by misleading graphs.

4. Center, variation, distribution, outliers, change in the characteristics of data over time. The time-series graph does the best job of giving us insight into the change in the characteristics of data over time.

5. The pulse rate of 36 beats per minute appears to be an outlier.



6. There do not appear to be any outliers.



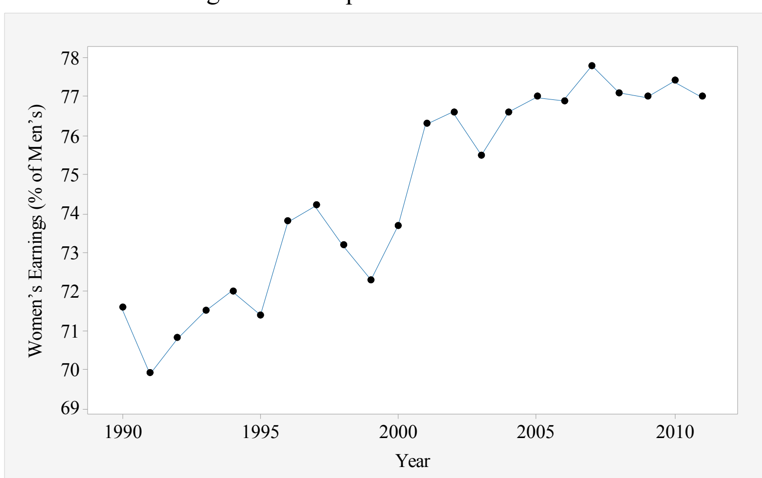
7. The data are arranged in order from lowest to highest, as 36, 56, 56, and so on.

3	6
4	
5	668
6	044666
7	6888
8	02468
9	4

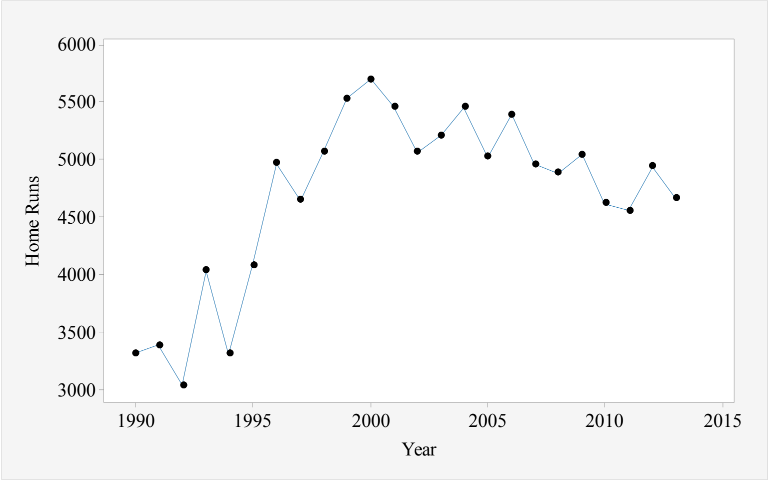
8. The two values closest to the middle are 72 mm Hg and 74 mm Hg.

6	0022468
7	0000246688
8	22468
9	00

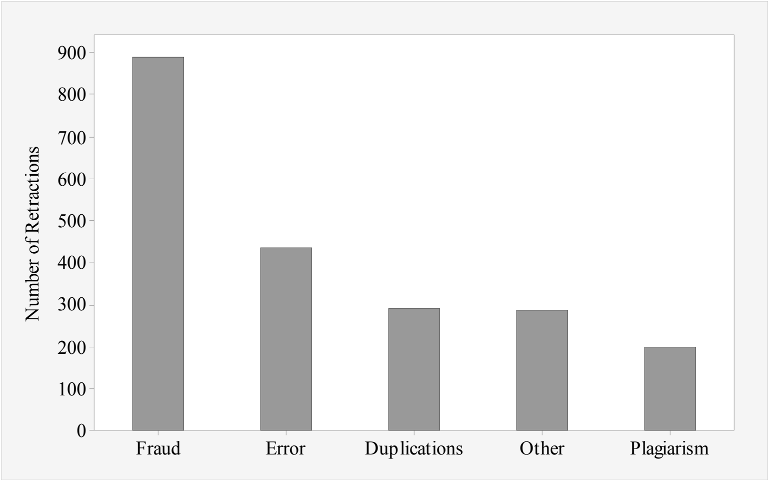
9. There is a gradual upward trend that appears to be leveling off in recent years. An upward trend would be helpful to women so that their earnings become equal to those of men.



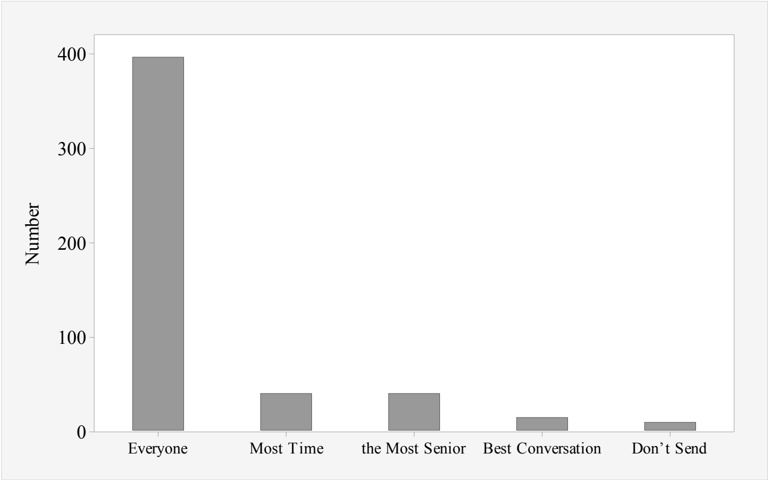
10. The numbers of home runs rose from 1990 to 2000, but after 2000 there has been a gradual decline.



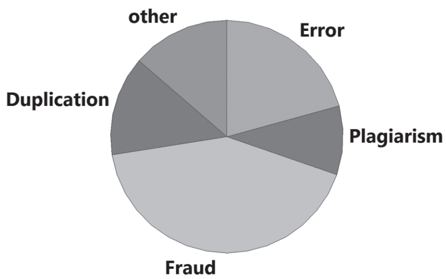
11. Misconduct includes fraud, duplication, and plagiarism, so it does appear to be a major factor.



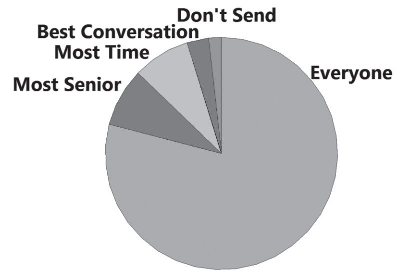
12. The overwhelming response was that thank-you notes should be sent to everyone who is met during a job interview. Given what is at stake, that seems like a wise strategy.



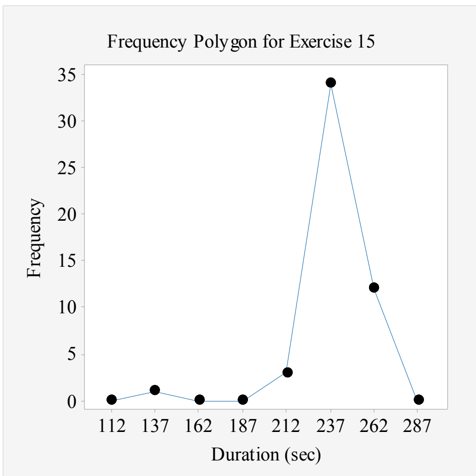
13.

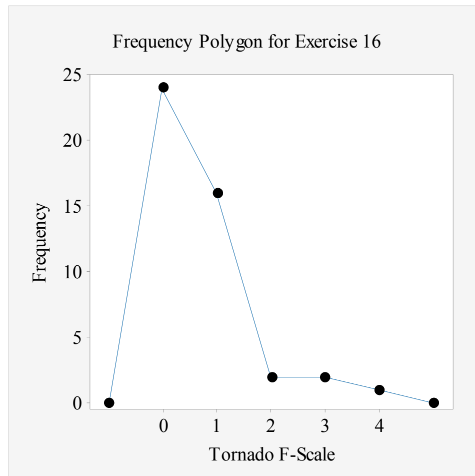


14.



15. The distribution appears to be skewed to the left (or negatively skewed).





16. The distribution appears to be skewed to the right (or positively skewed).

17. Because the vertical scale starts with a frequency of 200 instead of 0, the difference between the "no" and "yes" responses is greatly exaggerated. The graph makes it appear that about *five* times as many respondents said "no," when the ratio is actually a little less than 2.5 to 1.

18. The fare increased from $1 to $2.50, so it increased by a factor of 2.5. But when the larger bill is drawn so that the width is 2.5 times that of the smaller bill and the height is 2.5 times that of the smaller bill, the larger bill has an area that is 6.25 times that of the smaller bill (instead of being 2.5 times its size, as it should be). The illustration greatly exaggerates the increase in the fare.

19. The two costs are one-dimensional in nature, but the baby bottles are three-dimensional objects. The $4500 cost isn't even twice the $2600 cost, but the baby bottles make it appear that the larger cost is about five times the smaller cost.

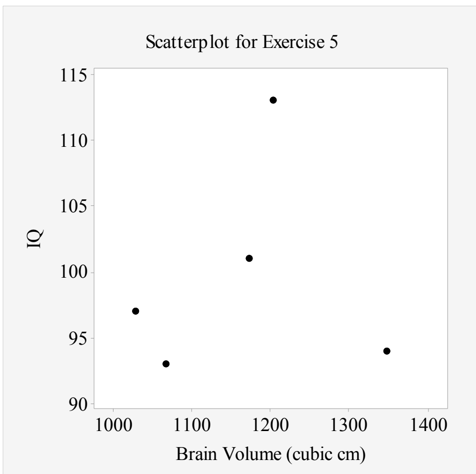
20. The graph is misleading because it depicts one-dimensional data with three-dimensional boxes. See the first and last boxes in the graph. Workers with advanced degrees have annual incomes that are roughly 3 times the incomes of those with no high school diplomas, but the graph exaggerates this difference by making it appear that workers with advanced degrees have incomes that are roughly 27 times the amounts for workers with no high school diploma.

21.

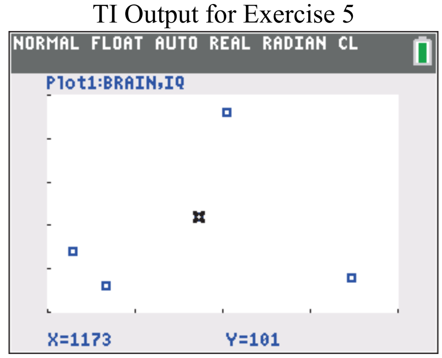
96.	59
97.	0001112333444
97.	55666666788888999
98.	000000000000022222334444444444444
98.	5555666666666666666777777888888899
99.	001244
99.	56

Section 2-4: Scatterplots, Correlation, and Regression

1. The term *linear* refers to a straight line, and r measures how well a scatterplot of the sample paired data fits a straight-line pattern.
2. No. Finding the presence of a statistical correlation between two variables does not justify any conclusion that one of the variables is a *cause* of the other.
3. A scatterplot is a graph of paired (x, y) quantitative data. It helps us by providing a visual image of the data plotted as points, and such an image is helpful in enabling us to see patterns in the data and to recognize that there may be a correlation between the two variables.
4. a. 1 c. 0
   b. 0 d. –1
5. There does not appear to be a linear correlation between brain volume and IQ score.

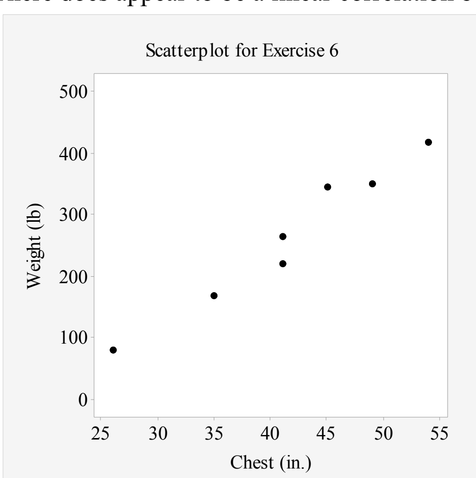


TI Output for Exercise 5

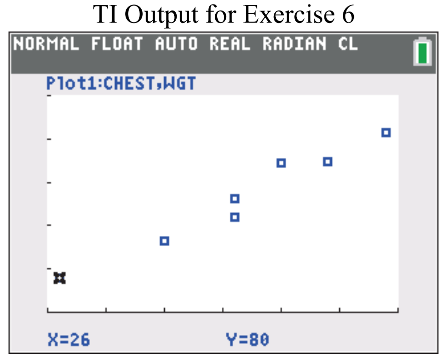


[1000, 1400] by [90, 115]

6. There does appear to be a linear correlation between chest sizes and weights of bears.

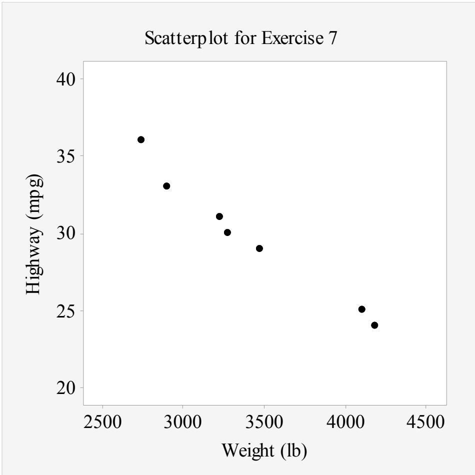


TI Output for Exercise 6

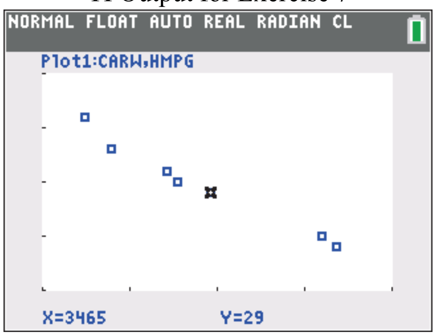


[25, 55] by [0, 500]

7. There does appear to be a linear correlation between weight and highway fuel consumption.

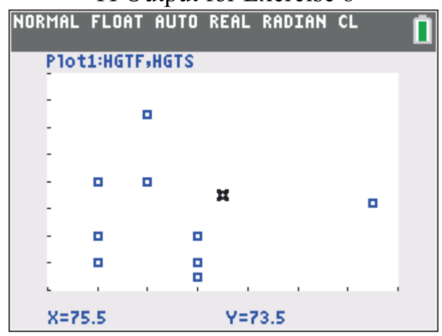


TI Output for Exercise 7



[2500, 4500] by [20, 40]

TI Output for Exercise 8



[72, 79] by [70, 78]

8. There does not appear to be a linear correlation between heights of fathers and the heights of their first sons.

9. With $n = 5$ pairs of data, the critical values are ±0.878. Because $r = 0.127$ is between –0.878 and 0.878, evidence is not sufficient to conclude that there is a linear correlation.

10. With $n = 7$ pairs of data, the critical values are ±0.754. Because $r = 0.980$ is in the right tail region beyond 0.754, there are sufficient data to conclude that there is a linear correlation.

11. With $n = 7$ pairs of data, the critical values are ±0.754. Because $r = 0.987$ is in the left tail region below –0.754, there are sufficient data to conclude that there is a linear correlation.

12. With $n = 10$ pairs of data, the critical values are ±0.632. Because $r = -0.017$ is between –0.632 and 0.632, evidence is not sufficient to conclude that there is a linear correlation.

13. Because the P-value is not small (such as 0.05 or less), there is a high chance (83.9% chance) of getting the sample results when there is no correlation, so evidence is not sufficient to conclude that there is a linear correlation.

14. Because the P-value is small (such as 0.05 or less), there is a small chance of getting the sample results when there is no correlation, so there is sufficient evidence to conclude that there is a linear correlation.

15. Because the P-value is small (such as 0.05 or less), there is a small chance of getting the sample results when there is no correlation, so there is sufficient evidence to conclude that there is a linear correlation.

16. Because the P-value is not small (such as 0.05 or less), there is a high chance (96.3% chance) of getting the sample results when there is no correlation, so the evidence is not sufficient to conclude that there is a linear correlation.

Chapter Quick Quiz

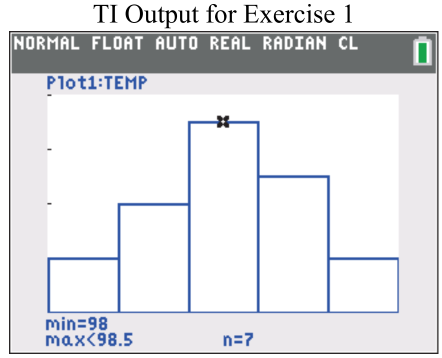
1. Class width: 3. It is not possible to identify the original data values.
2. Class boundaries: 17.5 and 20.5; Class limits: 18 and 20.
3. 40
4. 19 and 19
5. Pareto chart
6. histogram
7. scatterplot
8. No, the term "normal distribution" has a different meaning than the term "normal" that is used in ordinary speech. A normal distribution has a bell shape, but the randomly selected lottery digits will have a uniform or flat shape.
9. variation
10. The bars of the histogram start relatively low, increase to some maximum, and then decrease. Also, the histogram is symmetric, with the left half being roughly a mirror image of the right half.

Review Exercises

1.

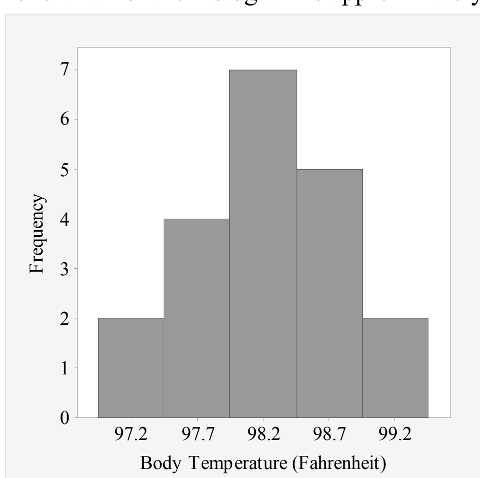
Temperature (°F)	Frequency
97.0–97.4	2
97.5–97.9	4
98.0–98.4	7
98.5–98.9	5
99.0–99.4	2

TI Output for Exercise 1

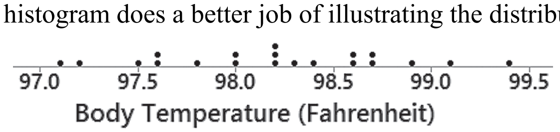


[97.0, 99.5] by [0, 8]

2. Yes, the data appear to be from a population with a normal distribution because the bars start low and reach a maximum, then decrease, and the left half of the histogram is approximately a mirror image of the right half.



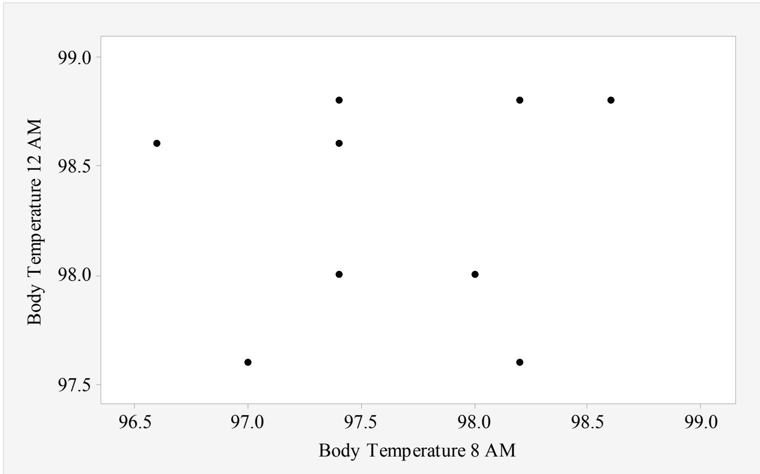
3. By using fewer classes, the histogram does a better job of illustrating the distribution.



4. There are no outliers.

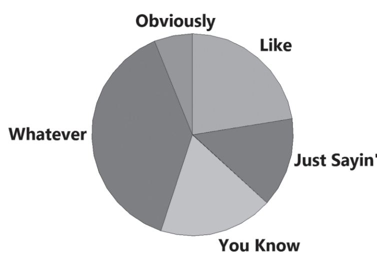
97.	125668
98.	002223466779
99.	14

5. No. There is no pattern suggesting that there is a relationship.

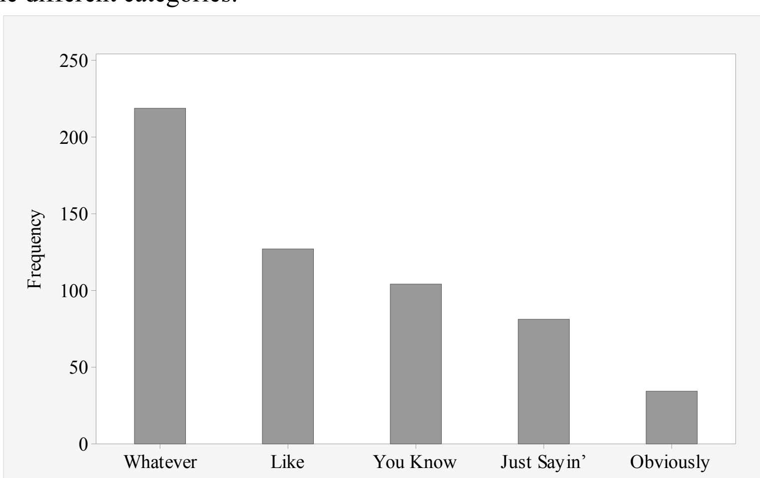


6. a. time-series graph
   b. scatterplot
   c. Pareto chart

7. A pie chart wastes ink on components that are not data; pie charts lack an appropriate scale; pie charts don't show relative sizes of different components as well as some other graphs, such as a Pareto chart.



8. The Pareto chart does a better job. It draws attention to the most annoying words or phrases and shows the relative sizes of the different categories.

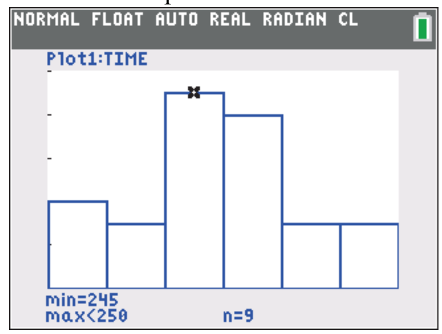


Cumulative Review Exercises

1.

Total Hours	Frequency
235–239	4
240–244	3
245–249	9
250–254	8
255–259	3
260–264	3

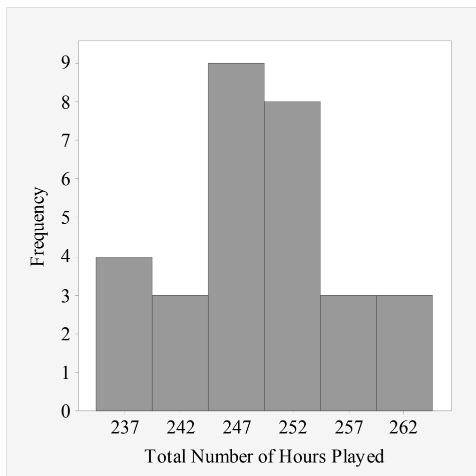
TI Output for Exercise 1



[235, 265] by [0, 10]

2. a. 235 hours and 239 hours
   b. 234.5 hours and 239.5 hours
   c. 237 hours

3. The distribution is closer to being a normal distribution than the others.



4. Start the vertical scale at a frequency of 2 instead of the frequency of 0.

5. Looking at the stemplot sideways, we can see that the distribution approximates a normal distribution.

23	6789
24	112555677889
25	00012233888
26	024

6. a. continuous
   b. quantitative
   c. ratio
   d. convenience sample
   e. sample

II. How Should Statistics Be Taught?

One of the most important points to be made in this *Insider's Guide* is the basic approach to teaching the introductory statistics course. Here are some important principles:

1. The introductory statistics course should be taught in a way that is fundamentally different from the approach used in traditional mathematics courses. Arithmetic computations or algebraic manipulations are not nearly as important as the ability to *understand* results and to be able to *interpret* results in a meaningful way.

2. The introductory statistics course focuses on *real applications* instead of abstractions.

3. Textbooks in the Triola Statistics Series are full of real data. Examples, exercises, and test questions should involve students with real data as much as possible. Fabricated data have little use in the introductory statistics course.

4. There should not be a high priority placed on covering as many different topics as possible. It is much better to cover fewer topics well than to cover many topics poorly.

The following pages identify the GAISE recommendations. The author comments about the GAISE recommendations are designed to clarify the above points.

GAISE Recommendations

GAISE is an acronym for "Guidelines for Assessment and Instruction in Statistics Education." These guidelines are recommendations from a project sponsored by the American Statistical Association (ASA). Here are six GAISE recommendations for the teaching of introductory statistics:

1. **Emphasize statistical literacy and develop statistical thinking.**

2. **Use real data.**

3. **Stress conceptual understanding rather than mere knowledge of procedures.**

4. **Foster active learning in the classroom.**

5. **Use technology for developing conceptual understanding and analyzing data.**

6. **Use assessments to improve and evaluate student learning.**

The author enthusiastically supports these recommendations, and much of the content of this *Insider's Guide* is devoted to implementation of these recommendations. Here are some comments about the six recommendations.

1. **Emphasize statistical literacy and develop statistical thinking.**
The importance of sound sampling techniques should be introduced early and often throughout the introductory statistics course. Part of "literacy" is understanding the meaning of terms such as *simple random sample* and *voluntary response sample*. Statistical thinking is used when a student recognizes that results obtained from a poorly selected sample might be results without any real validity. For example, newspapers, magazines, television shows, and Internet sites often conduct surveys by asking people to respond to some question. However, the responses constitute a voluntary response sample, and students should know that any conclusions based on such a sample do not apply to the larger population. This is one simple example of the type of critical thinking that should be fostered throughout the course.

In teaching the introductory statistics course, it is not important to memorize formulas or the detailed mechanics of statistical methods. It is not important to be able to reproduce the formula for the standard deviation s, and it is not so important to be able to do the arithmetic required for manually computing values of standard deviations. Instead, it is important to *understand* what the standard deviation s measures. On a very basic level, it is important for students to know quite well that s is a measure of *variation*. It is *really* important that students develop an ability to *understand* and *interpret* values of the standard deviation s. The empirical rule and Chebyshev's theorem are commonly presented as tools that help students understand and interpret s, but the author recommends skipping those two topics and focusing instead on the *range rule of thumb* presented in the book. It is easy to apply, and students generally understand it quite well, so it becomes a very effective tool that can help students understand and interpret values of standard deviations. This topic will be discussed further when measures of variation are discussed later in this guide. But this topic is excellent for making the point that we should emphasize statistical literacy and develop statistical thinking.

When teaching an introductory calculus course, the author might give a test question that asks students to write the definition of the derivative of a function $f(x)$, and he might ask students to compute the derivative of $f(x) = x^2$ while showing all of the steps involved. Calculus students should know the definition of the derivative and they should be able to apply it. However, the author would never ask statistics students to write the formula for the standard deviation or to calculate the standard deviation of a list of values while showing all work. Instead, the author prefers to ask questions that test *understanding*. Here are examples of good and bad test questions:

Bad test question: Write the formulas for the mean and standard deviation *s,* then compute the mean and standard deviation of the values 23.7, 11.2, 43.5, 77.2, 49.0, 27.3, and show all work.

Good test question: Listed below are weights (in grams) of newly minted quarters. (a) Find the mean. (b) Find the standard deviation. (c) In the context of the given weights, is a weight of 5.23 g *significantly low*? Explain your choice. (d) What is an adverse consequence of minting quarters with weights that vary too much?

5.71 5.71 5.59 5.61 5.63

When students find the mean $\bar{x}$ = 5.650 g and standard deviation of s = 0.057 g, they should be encouraged to use some technology, such as a TI-83 Plus or TI-84 Plus calculator. There is little to be gained by requiring that such statistics be calculated manually. A good answer to part (c) of the preceding question is the statement that yes, a weight of 5.23 g would be significantly low because it is more than two standard deviations away from the mean. One of several good answers to part (d) would be a statement that if weights of minted quarters vary too much, vending machines will reject too many valid coins. Part (d) is designed to emphasize the point that methods of statistics have real, important, and meaningful applications instead of being abstract concepts that might not have any real applications.

2. Use real data

George Cobb is a leader in statistics education. He wrote an article about evaluating introductory statistics textbooks (see "Introductory Textbooks: A Framework for evaluation", *Journal of the American Statistical Association*, Vol. 82, No. 397) and he included the following statement:

"Are the Data Sets Real or Fake? Not that many years ago, all it took was this first question to dispatch most books to the morgue. Fortunately, that is changing. It is true that there are still books on the market whose examples have been bled white of vital detail, but it is now easier to shun them. I hope that soon we will have seen the last of the infamous XYZ Corporation and Hospitals A, B, C, ..."

In the 13th edition of *Elementary Statistics*, 94% of the examples involve real data, and 92% of the exercises involve real data. With real data, students see how statistical concepts have meaningful applications. It is very likely that they will encounter data from the discipline that they might be considering as a major.

3. Stress conceptual understanding rather than mere knowledge of procedures

A good illustration of this point can be seen in the data from eruptions of the Old Faithful geyser and data from actual low temperatures and forecast temperatures:

Duration (sec)	240	120	178	234	235	269	255	220
Interval After (min)	92	65	72	94	83	94	101	87

Actual low (^{o}F)	54	54	55	60	64
Low forecast five days earlier (^{o}F)	56	57	59	56	64

When discussing correlation/regression, we might present the top table and ask if there is a correlation between the duration of an eruption and the time interval after the eruption to the next eruption. When discussing matched data, we might present the bottom table given above, and we might ask if the differences between the actual and forecast temperatures are from a population with a mean of 0. But instead of focusing too much on the details of the computations involved, we should stress the fundamental difference between the two sets of data summarized in the preceding tables. Students should learn how to ask the best questions. Given the top Old Faithful table, students should see that the issue is one of a *relationship* between the two variables. Given the bottom temperature table, students should see that a key element is the list of *differences* between the actual and forecast temperatures, and a mean difference equal to zero is evidence that the forecast temperatures are accurate. It's not the structure of paired data that determines the method that is most appropriate; it is the *context* of the data.

4. Foster active learning in the classroom

Here is a saying that is so true when considered in the context of teaching an introductory statistics course:

Tell me something, and I will forget.
Show me and I will remember.
Involve me, and I will learn.

If you want your students to have a learning experience that will affect them for their entire lives, *involve* them with active learning. *Elementary Statistics,* 13th Edition, has Cooperative Group Activities at the end of each chapter, except for Chapter 15.

Some statistics professors believe that the entire course should be based on activities, and some other statistics professors do not include any activities at all. Somewhere between these extremes is a balance that allows active involvement along with enough time for teaching concepts using traditional methods.

Recommendation: If you do no activities at all, begin with just one or two activities to see how well they work. Then, assuming that all goes well, include more activities in future courses.

5. Use technology for developing conceptual understanding and analyzing data

Many statistics professors teach an effective course by allowing students to use any one of a variety of different scientific calculators. The author recommends that a specific technology be used. Triola statistics books include displays from Statdisk, Minitab, Excel, the TI-83/84 Plus calculator, and StatCrunch. There are also supplements for SPSS and SAS.

The author's personal preference is to require that each student have a TI-83 Plus or TI-84 Plus calculator, and that each student also do several software projects using Statdisk. However, choosing a technology to be used for an introductory statistics course is a complex decision that must take several factors into account. Some colleges have adopted a decision to use Excel because so many students use Excel in their work after graduation. Some colleges avoid Excel because its statistics functions are not as good as they should be. Some colleges use Minitab, and the latest release includes features that make it a perfectly good choice. Some statistics professors prefer to require TI-83/84 Plus calculators because they can do so much statistical number crunching and they can be used in class and on tests. Some statistics professors would like to require TI-83/84 Plus calculators, but are reluctant to do so because of the calculator's cost. The author had that same concern the first time that he required those calculators, so he announced that any student could sell him their calculator at the end of the course. At the end of that semester, *no* students wanted to give up their calculators. Their desire to keep their calculators instead of turning them in for cash was a strong indication about how they perceived the usefulness of those calculators.

Statdisk Statdisk is a free and easy-to-use software package designed specifically for the Triola Statistics Series textbooks. The latest version of Statdisk is one that the author is proud to have as a major and important supplement. Because Statdisk can do almost all of the functions described in the textbook, it can be used as the technology in the introductory statistics course. If another technology, such as Excel or SPSS, is used as the primary technology, it would be really helpful to have students use Statdisk as a supplement to the main technology being used. By getting results from Statdisk along with results from another technology, students are more likely to confirm that their results are correct.

Technology for New Approaches While the technology has the ability to do the statistics number crunching, it should also be used to explore concepts and new approaches. When considering the effects of an outlier, for example, a hypothesis test could be conducted with the outlier included and again with the outlier excluded. Probability can be better studied with simulations. Bootstrap resampling techniques can sometimes be used when traditional methods should not be used. For ideas about how to include technology, see the Technology Project at the end of each chapter in the Triola textbook

6. Use assessments to improve and evaluate student learning

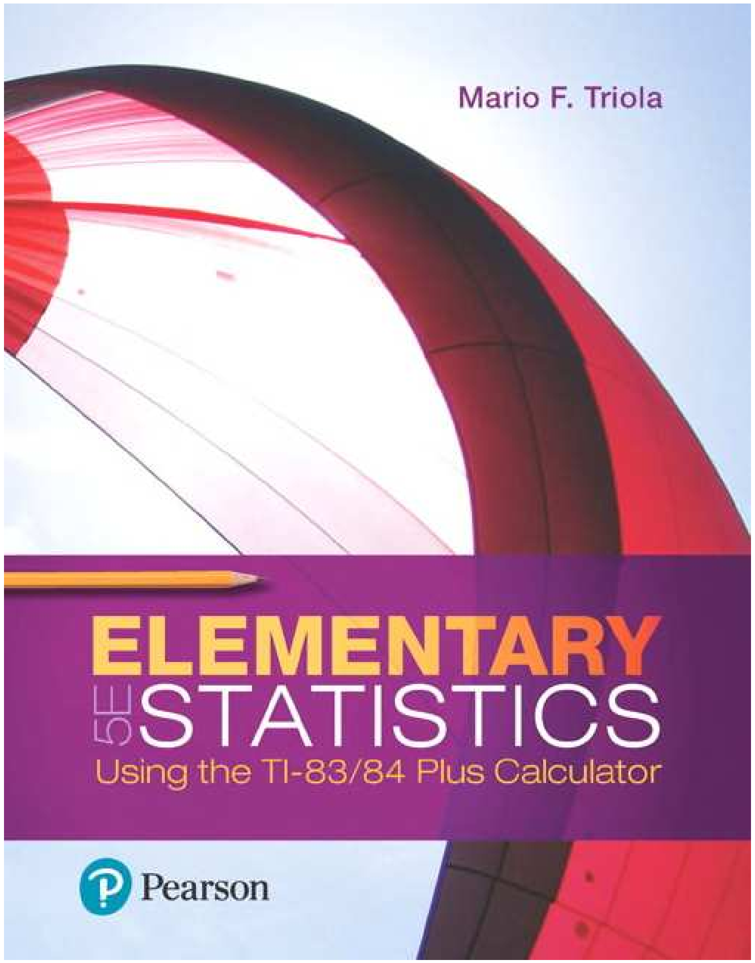
Traditional tests and quizzes are one important method of assessment, but there are others. The author favors the use of activities and at least one major project. The author favors a capstone group project conducted near the end of the course. Students can work together in groups of four (more or less), and each group should conduct a project that involves the planning of an experiment or a method for collecting data in an observational study. After collecting original data, the group will make an inference by using the methods learned in the course.

A group presentation should involve each member speaking for at least one or two minutes. A computer display should also be included, along with a brief written report. Assessment is an important component of such a project. How do you assess the work of individual members that participate in a group project? Here is one method that the author found to be effective: Survey each group member and ask him or her to assess the work done by the other group members. For example, ask each group member to submit a separate form for each of the other group members, and that form should include an assessment of the other team members' work, such as "was a major contributor to the project," "did an average amount of work on the project," "did some but little work on the project," or "did not participate in any meaningful way." Students are quite honest about the work of their peers, and they are generally quite satisfied with this process of assessment.

The author favors four or five tests given during the semester, along with a comprehensive final examination. Activities and projects should also be part of the assessment plan.

Elementary Statistics Using the TI-83/84 Plus Calculator

Fifth Edition



Chapter 2

Exploring Data with Tables and Graphs

Exploring Data with Tables and Graphs

2-1 Frequency Distributions for Organizing and Summarizing Data

2-2 Histograms

2-3 Graphs that Enlighten and Graphs that Deceive

2-4 Scatterplots, Correlation, and Regression

Key Concept

When working with large data sets, a **frequency distribution** (or **frequency table**) is often helpful in organizing and summarizing data. A frequency distribution helps us to understand the nature of the **distribution** of a data set.

Frequency Distribution

- Frequency Distribution (or Frequency Table)
  - Shows how data are partitioned among several categories (or **classes**) by listing the categories along with the number (frequency) of data values in each of them.

Definitions (1 of 2)

- Lower class limits
  - The smallest numbers that can belong to each of the different classes
- Upper class limits
  - The largest numbers that can belong to each of the different classes
- Class boundaries
  - The numbers used to separate the classes, but without the gaps created by class limits

Definitions (2 of 2)

- Class midpoints
  – The values in the middle of the classes Each class midpoint can be found by adding the lower class limit to the upper class limit and dividing the sum by 2.
- Class width
  – The difference between two consecutive lower class limits in a frequency distribution

Procedure for Constructing a Frequency Distribution (1 of 2)

1. Select the number of classes, usually between 5 and 20.
2. Calculate the class width.

$$\text{Class width} \approx \frac{(\text{maximum data value}) - (\text{minimum data value})}{\text{number of classes}}$$

Round this result to get a convenient number. (It’s usually best to round **up**.)

Procedure for Constructing a Frequency Distribution (2 of 2)

3. Choose the value for the first lower class limit by using either the minimum value or a convenient value below the minimum.

4. Using the first lower class limit and class width, list the other lower class limits.

5. List the lower class limits in a vertical column and then determine and enter the upper class limits.

6. Take each individual data value and put a tally mark in the appropriate class. Add the tally marks to get the frequency.

Example: McDonald's Lunch Service Times (1 of 5)

Using the McDonald's lunch service times in the first table, follow the procedure shown on the next slide to construct the frequency distribution shown in the second table. Use five classes.

Drive-through Service Times (seconds) for McDonald's Lunches

107	139	197	209	281	254	163	150	127	308	206	187	169	83	127	133	140
143	130	144	91	113	153	255	252	200	117	167	148	184	123	153	155	154
100	117	101	138	186	196	146	90	144	119	135	151	197	171	190	169	

McDonald's Lunch Drive-Through Service Times

Time (Seconds)	Frequency
75-124	11
125-174	24
175-224	10
225-274	3
275-324	2

Example: McDonald's Lunch Service Times (2 of 5)

Step 1: Select 5 as the number of desired classes.

Step 2: Calculate the class width as shown below. Note that we round 45 up to 50,which is a more convenient number.

$$\text{Class width} \approx \frac{(\text{maximum data value}) - (\text{minimum data value})}{\text{number of classes}}$$

$$= \frac{308 - 83}{5} = 45 \approx 50 \text{ (rounded up to a more convenient number)}$$

Example: McDonald's Lunch Service Times (3 of 5)

Step 3: The minimum data value is 83, which is not a very convenient starting point, so go to a value below 83 and select the more convenient value of 75 as the first lower class limit.

Step 4: Add the class width of 50 to the starting value of 75 to get the second lower class limit of 125. Continue to add the class width of 50 until we have five lower class limits. The lower class limits are therefore 75, 125, 175, 225, and 275.

Example: McDonald's Lunch Service Times (4 of 5)

Step 5: List the lower class limits vertically, as shown below. From this list, we identify the corresponding upper class limits as 124, 174, 224, 274, and 324.

75-
125-
175-
225-
275-

Example: McDonald’s Lunch Service Times (5 of 5)

Step 6: Enter a tally mark for each data value in the appropriate class. Then add the tally marks to find the frequencies shown in the table.

Time (Seconds)	Frequency
75-124	11
125-174	24
175-224	10
225-274	3
275-324	2

Relative Frequency Distribution (1 of 2)

- Relative Frequency Distribution or Percentage Frequency Distribution
  - Each class frequency is replaced by a relative frequency (or proportion) or a percentage.

$$\text{Relative frequency for a class} = \frac{\text{frequency for a class}}{\text{sum of all frequencies}}$$

$$\text{Percentage for a class} = \frac{\text{frequency for a class}}{\text{sum of all frequencies}} \times 100\%$$

Relative Frequency Distribution (2 of 2)

- Relative Frequency Distribution or Percentage Frequency Distribution
  – Each class frequency is replaced by a relative frequency (or proportion) or a percentage.

The sum of the percentages in a relative frequency distribution must be very close to 100% (with a little wiggle room for rounding errors).

Cumulative Frequency Distribution

- Cumulative Frequency Distribution
  - The frequency for each class is the sum of the frequencies for that class and all previous classes.

Cumulative Frequency Distribution of McDonald's Lunch Service Times

Time (Seconds)	Cumulative Frequency
Less than 125	11
Less than 175	35
Less than 225	45
Less than 275	48
Less than 325	50

Critical Thinking: Using Frequency Distributions to Understand Data

In statistics we are often interested in determining whether the data have a **normal distribution**.

1. The frequencies start low, then increase to one or two high frequencies, and then decrease to a low frequency.
2. The distribution is approximately symmetric. Frequencies preceding the maximum frequency should be roughly a mirror image of those that follow the maximum frequency.

Gaps

- The presence of gaps can show that the data are from two or more different populations.
- However, the converse is not true, because data from different populations do not necessarily result in gaps.

Example: Exploring Data: What Does a Gap Tell Us? (1 of 2)

The table shown is a frequency distribution of the weights (grams) of randomly selected pennies.

Weight (grams) of Penny	Frequency
2.40-2.49	18
2.50-2.59	19
2.60-2.69	0
2.70-2.79	0
2.80-2.89	0
2.90-2.99	2
3.00-3.09	25
3.10-3.19	8

Example: Exploring Data: What Does a Gap Tell Us? (2 of 2)

- Examination of the frequencies reveals a large **gap** between the lightest pennies and the heaviest pennies.
- This suggests that we have two different populations:
  - Pennies made before 1983 are 95% copper and 5% zinc.
  - Pennies made after 1983 are 2.5% copper and 97.5% zinc.

Comparisons

Combining two or more relative frequency distributions in one table makes comparisons of data much easier.

Example: Comparing McDonald's and Dunkin' Donuts (1 of 2)

The table shows the relative frequency distributions for the drive-through lunch service times (seconds) for McDonald's and Dunkin' Donuts.

Time (seconds)	McDonald's	Dunkin' Donuts
25-74		22%
75-124	22%	44%
125-174	48%	28%
175-224	20%	6%
225-274	6%	
275-324	4%	

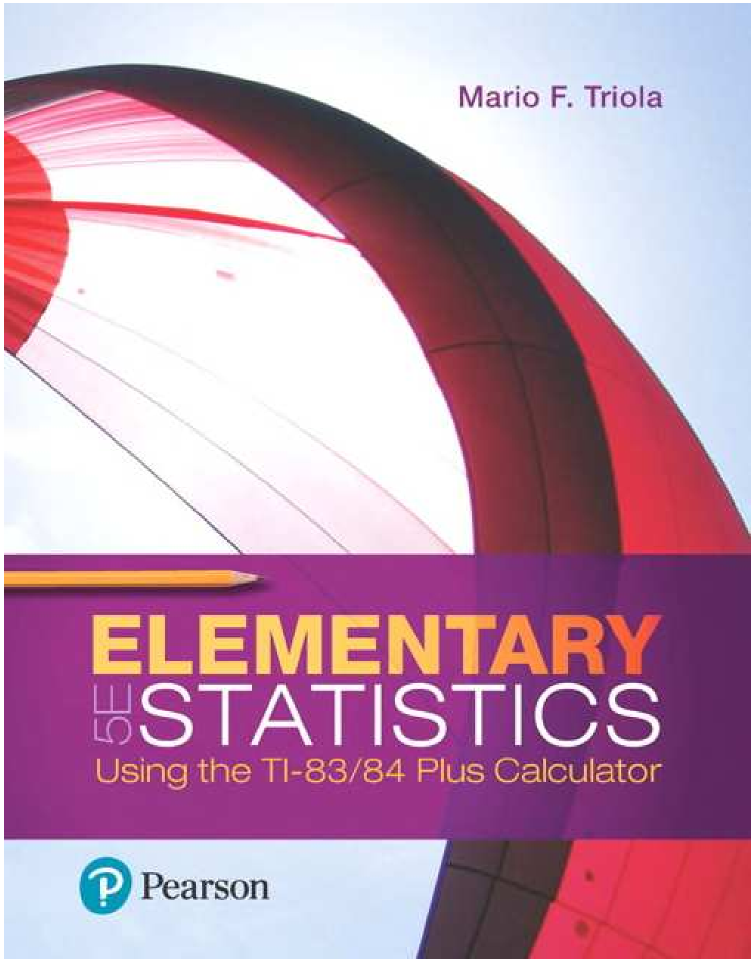
Example: Comparing McDonald's and Dunkin' Donuts (2 of 2)

Time (seconds)	McDonald's	Dunkin' Donuts
25-74		22%
75-124	22%	44%
125-174	48%	28%
175-224	20%	6%
225-274	6%	
275-324	4%	

- Because of the dramatic differences in their menus, we might expect the service times to be very different.
- By comparing the relative frequencies, we see that there are major differences. The Dunkin' Donuts service times appear to be lower than those at McDonald's.

Elementary Statistics Using the TI-83/84 Plus Calculator

Fifth Edition



Chapter 2

Exploring Data with Tables and Graphs

Exploring Data with Tables and Graphs

2-1 Frequency Distributions for Organizing and Summarizing Data

2-2 Histograms

2-3 Graphs that Enlighten and Graphs that Deceive

2-4 Scatterplots, Correlation, and Regression

Key Concept

While a frequency distribution is a useful tool for summarizing data and investigating the distribution of data, an even better tool is a **histogram**, which is a graph that is easier to interpret than a table of numbers.

Histogram

- **Histogram**
  – A graph consisting of bars of equal width drawn adjacent to each other (unless there are gaps in the data)

    The horizontal scale represents classes of quantitative data values, and the vertical scale represents frequencies. The heights of the bars correspond to frequency values.

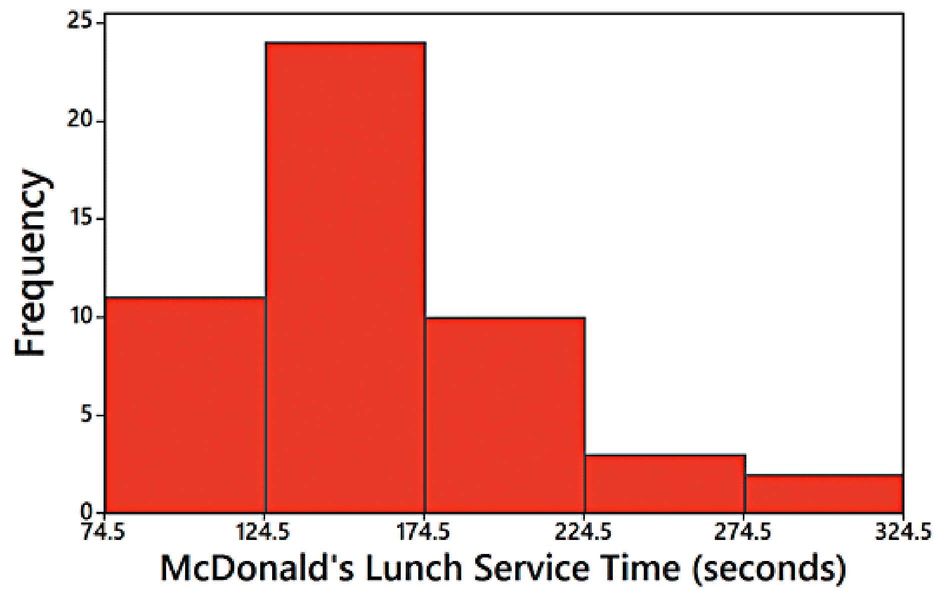
Important Uses of a Histogram

- Visually displays the shape of the **distribution** of the data
- Shows the location of the **center** of the data
- Shows the **spread** of the data
- Identifies **outliers**

Relative Frequency Histogram

- **Relative Frequency Histogram**
  - It has the same shape and horizontal scale as a histogram, but the vertical scale is marked with relative frequencies instead of actual frequencies.

Time (seconds)	Frequency
75-124	11
125-174	24
175-224	10
225-274	3
275-324	2

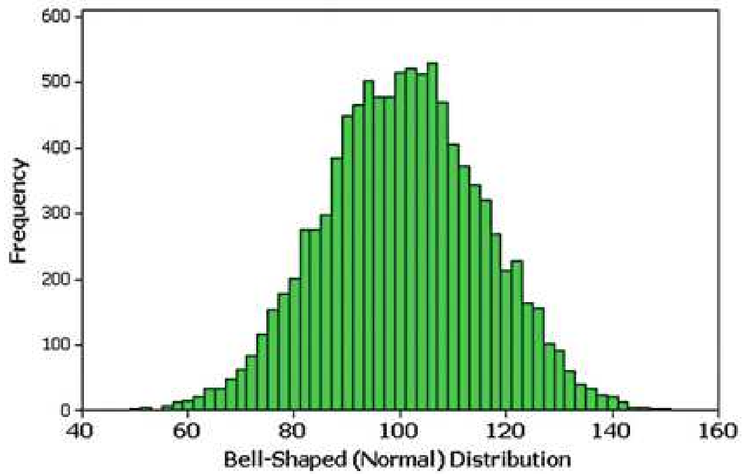


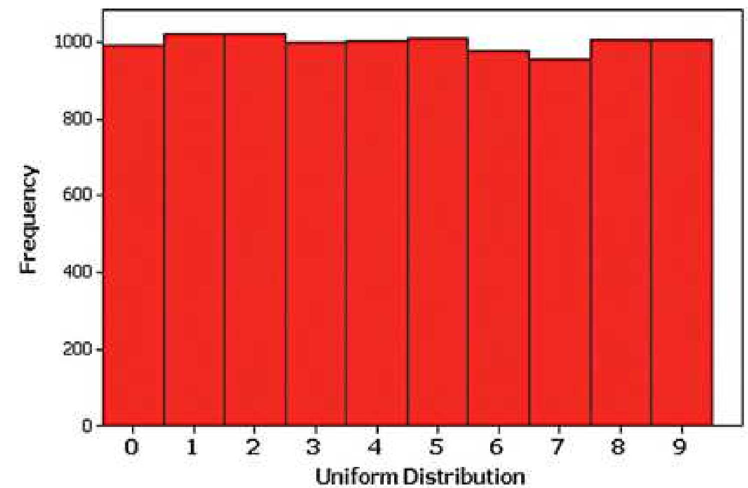
Critical Thinking Interpreting Histograms

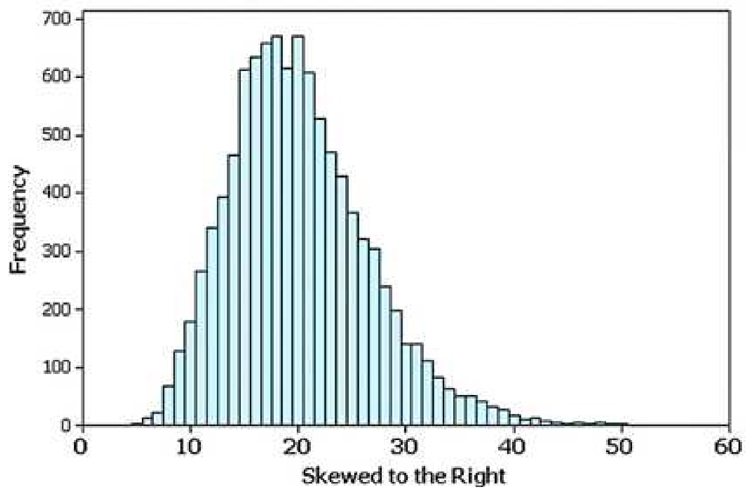
Explore the data by analyzing the histogram to see what can be learned about "**CVDOT**":

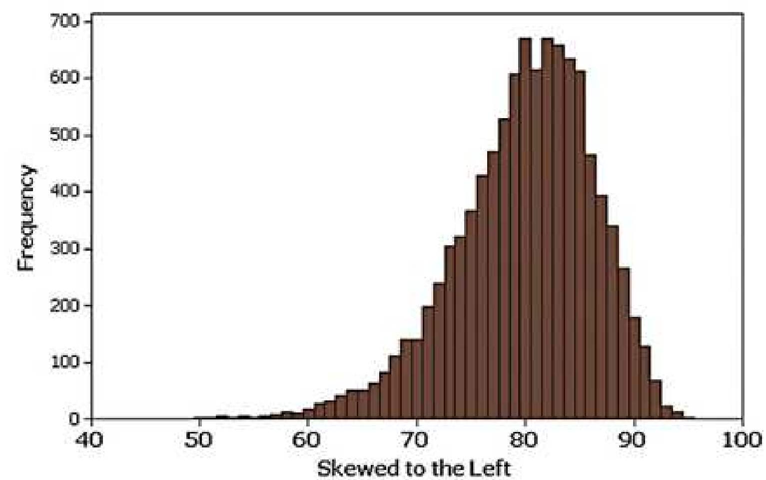
- the **C**enter of the data,
- the **V**ariation,
- the shape of the **D**istribution,
- whether there are any **O**utliers,
- and **T**ime.

Common Distribution Shapes





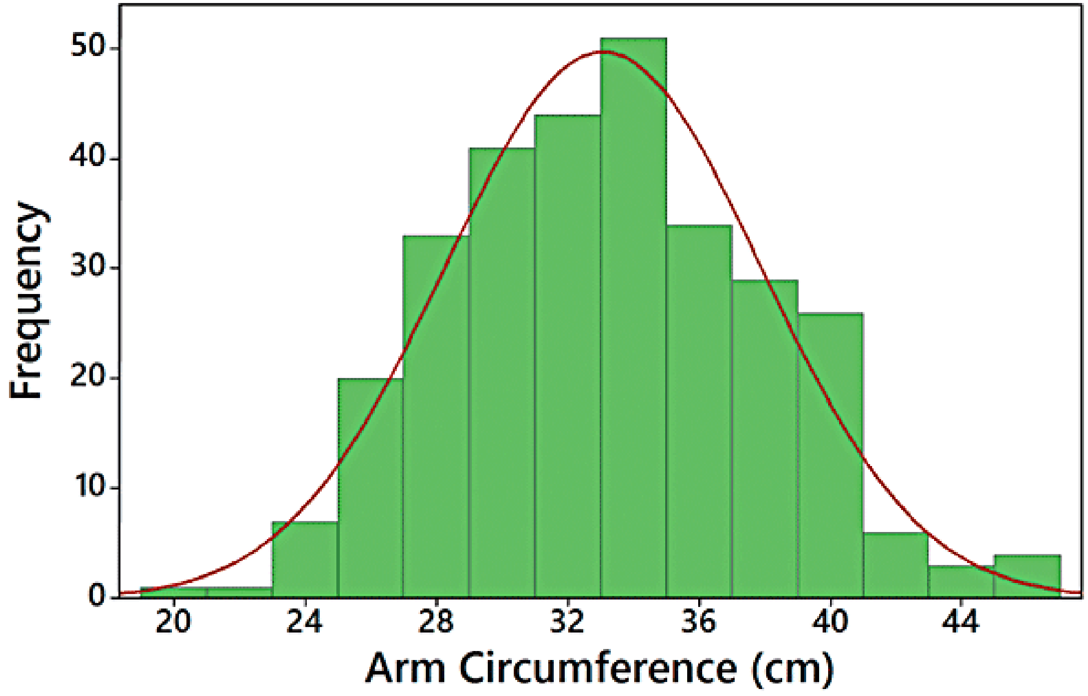




Pearson

Normal Distribution

Because this histogram is roughly bell-shaped, we say that the data have a **normal distribution**.

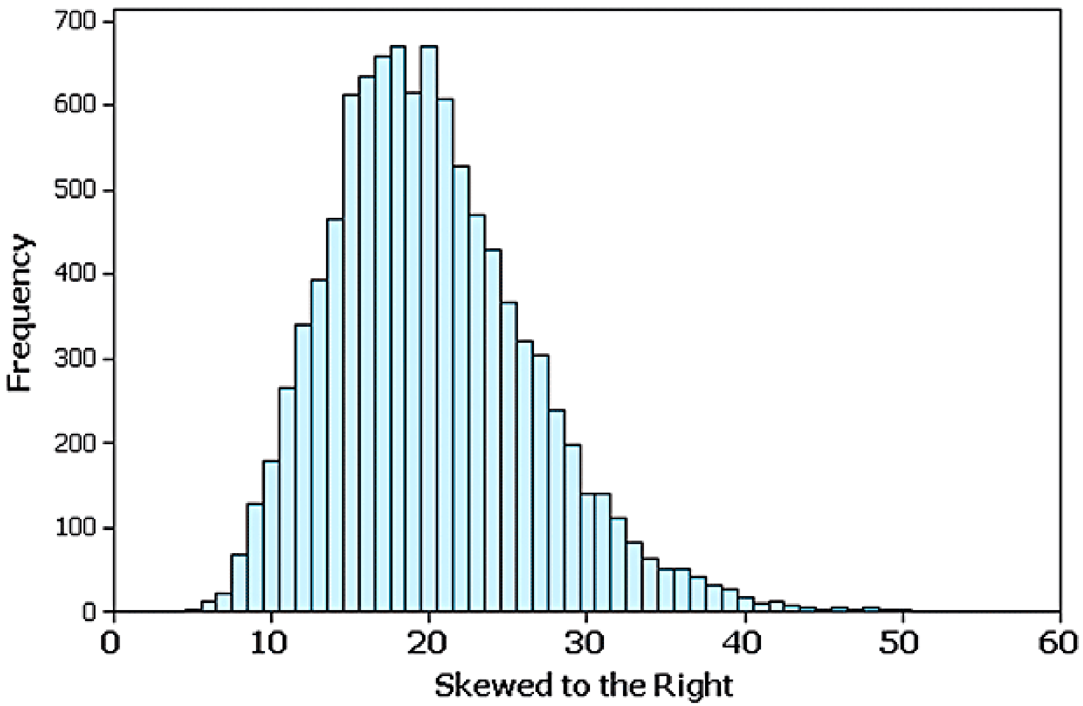


Skewness (1 of 3)

- **Skewness**
  - A distribution of data is **skewed** if it is not symmetric and extends more to one side than to the other.

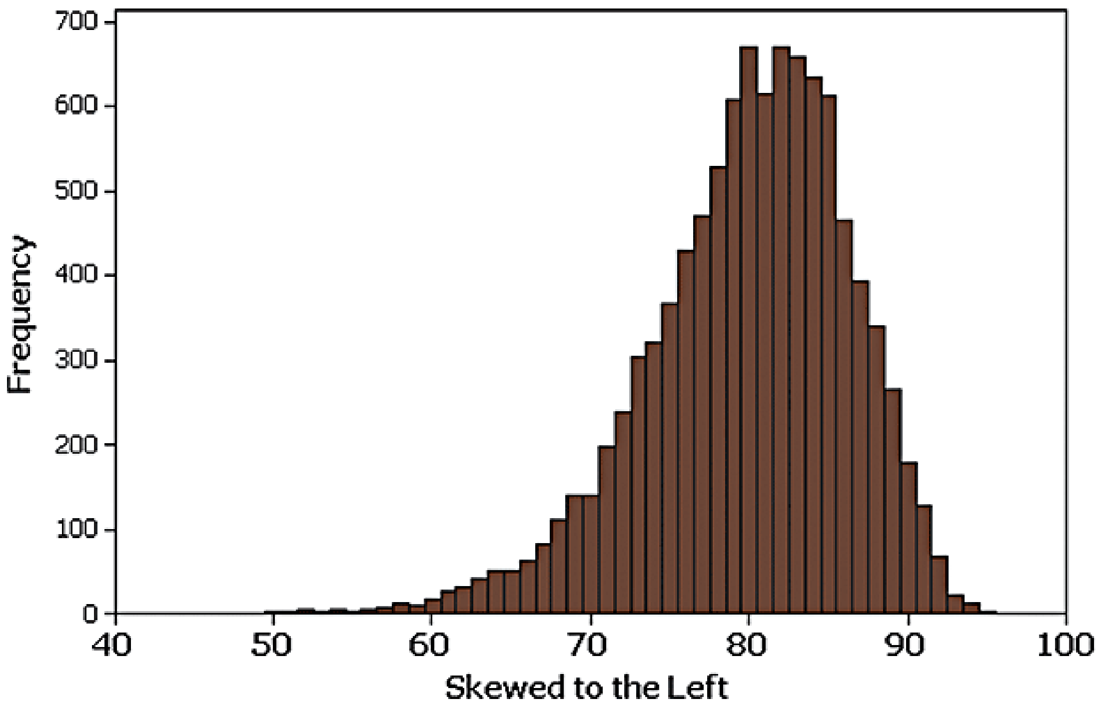
Skewness (2 of 3)

Data **skewed to the right (positively skewed)** have a longer right tail.



Skewness (3 of 3)

Data **skewed to the left (negative skewed)** have a longer left tail.



Assessing Normality with Normal Quantile Plots (1 of 5)

Criteria for Assessing Normality with a Normal Quantile Plot

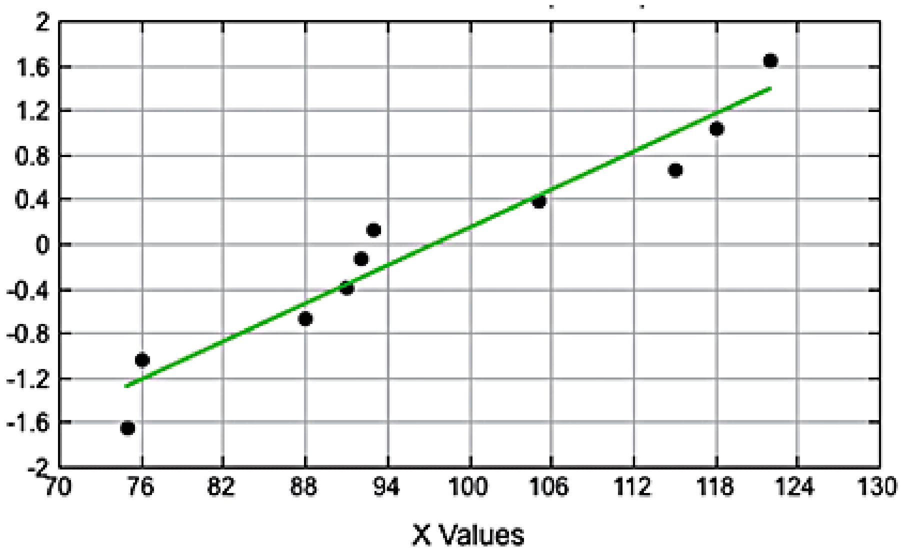
- **Normal Distribution:** The pattern of the points in the normal quantile plot is reasonably close to a straight line, and the points do not show some systematic pattern that is not a straight-line pattern.

Assessing Normality with Normal Quantile Plots (2 of 5)

Criteria for Assessing Normality with a Normal Quantile Plot

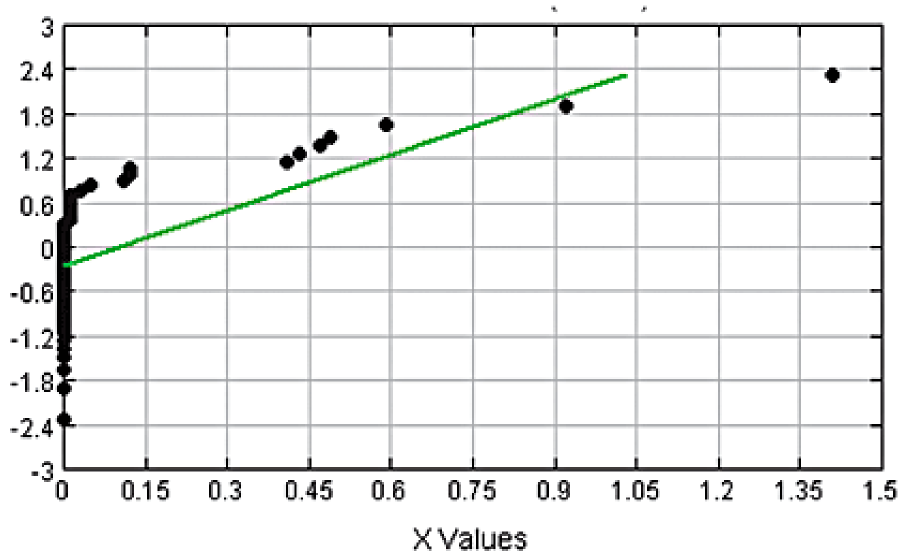
- **Not a Normal Distribution:** The population distribution is **not** normal if the normal quantile plot has either or both of these two conditions:
  - The points do not lie reasonably close to a straight-line pattern.
  - The points show some systematic pattern that is not a straight-line pattern.

Assessing Normality with Normal Quantile Plots (3 of 5)



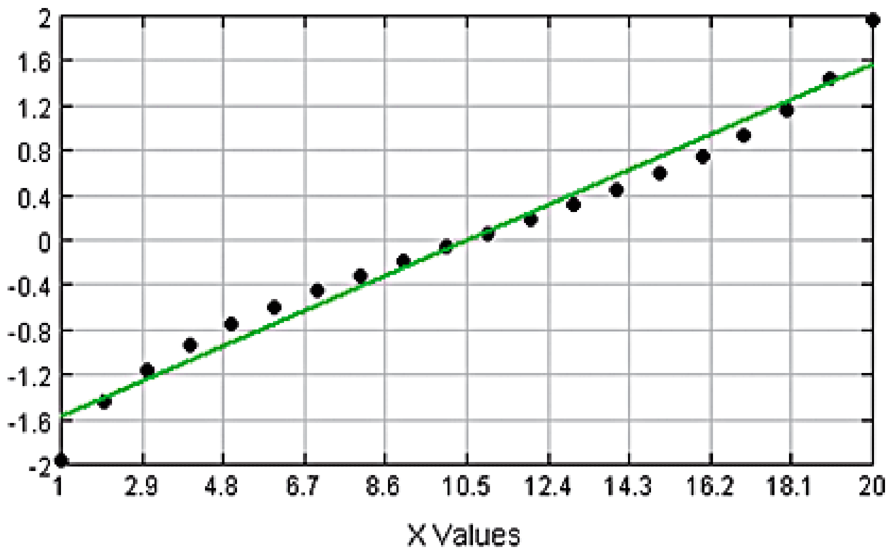
Normal Distribution: The points are reasonably close to a straight-line pattern, and there is no other systematic pattern that is not a straight-line pattern.

Assessing Normality with Normal Quantile Plots (4 of 5)



Not a Normal Distribution: The points do not lie reasonably close to a straight line.

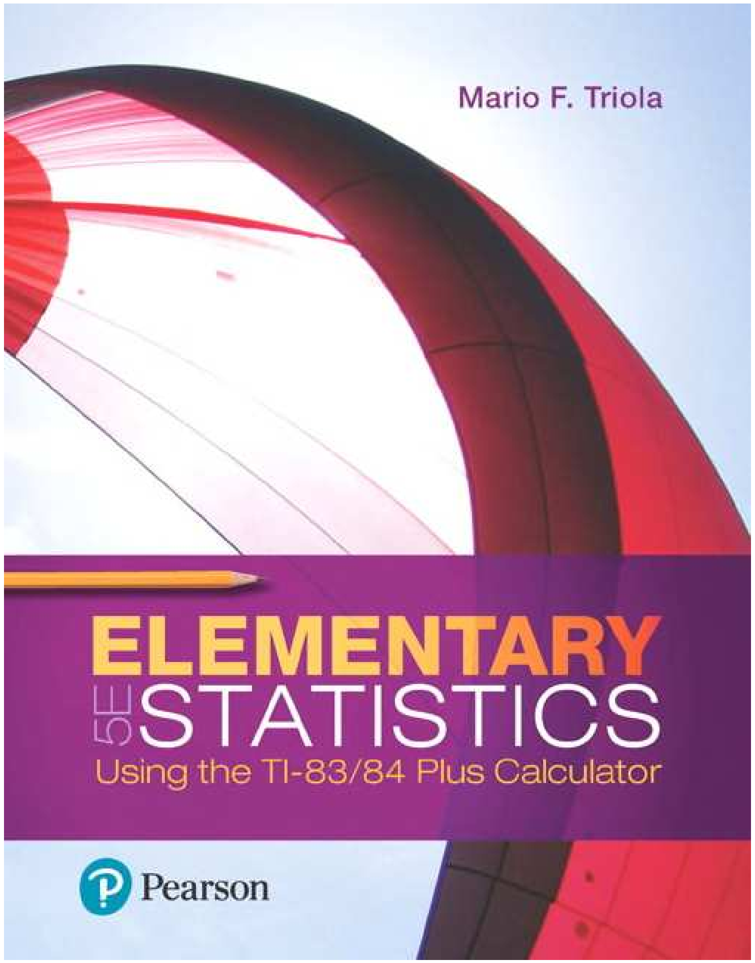
Assessing Normality with Normal Quantile Plots (5 of 5)



Not a Normal Distribution: The points show a systematic pattern that is not a straight-line pattern.

Elementary Statistics Using the TI-83/84 Plus Calculator

Fifth Edition



Chapter 2

Exploring Data with Tables and Graphs

Exploring Data with Tables and Graphs

2-1 Frequency Distributions for Organizing and Summarizing Data

2-2 Histograms

2-3 Graphs that Enlighten and Graphs that Deceive

2-4 Scatterplots, Correlation, and Regression

Key Concept

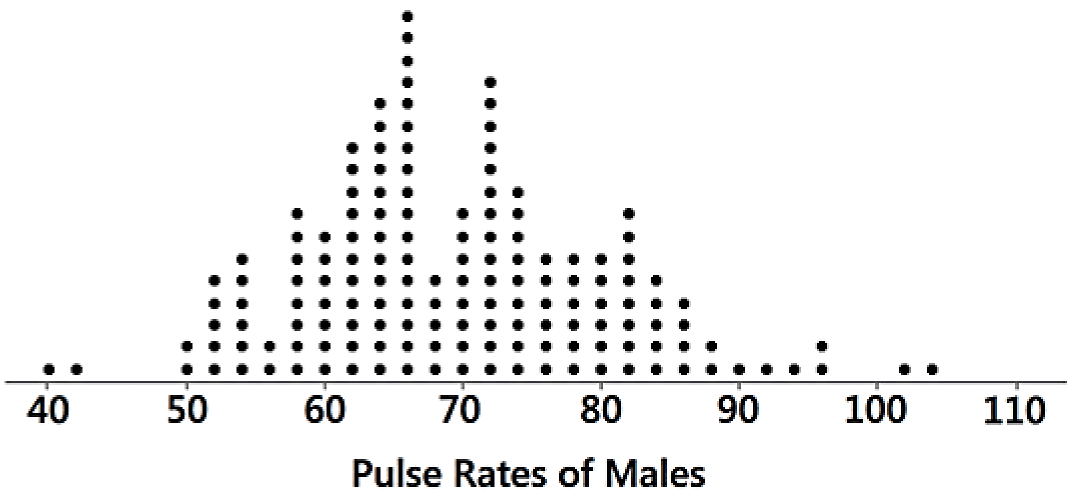
Introduce other common graphs that foster understanding of data.

Discuss some graphs that are deceptive because they create impressions about data that are somehow misleading or wrong.

Technology now provides us with powerful tools for generating a wide variety of graphs.

Graphs that Enlighten: Dotplots (1 of 2)

- **Dotplots**
  - A graph of **quantitative** data in which each data value is plotted as a point (or dot) above a horizontal scale of values. Dots representing equal values are stacked.



Graphs that Enlighten: Dotplots (2 of 2)

- **Dotplots**
  - **Features of a Dotplot**
    - Displays the shape of distribution of data.
    - It is usually possible to recreate the original list of data values.

Stemplots (1 of 2)

- **Stemplots (or stem-and-leaf plot)**
  - Represents **quantitative** data by separating each value into two parts: the stem (such as the leftmost digit) and the leaf (such as the rightmost digit).

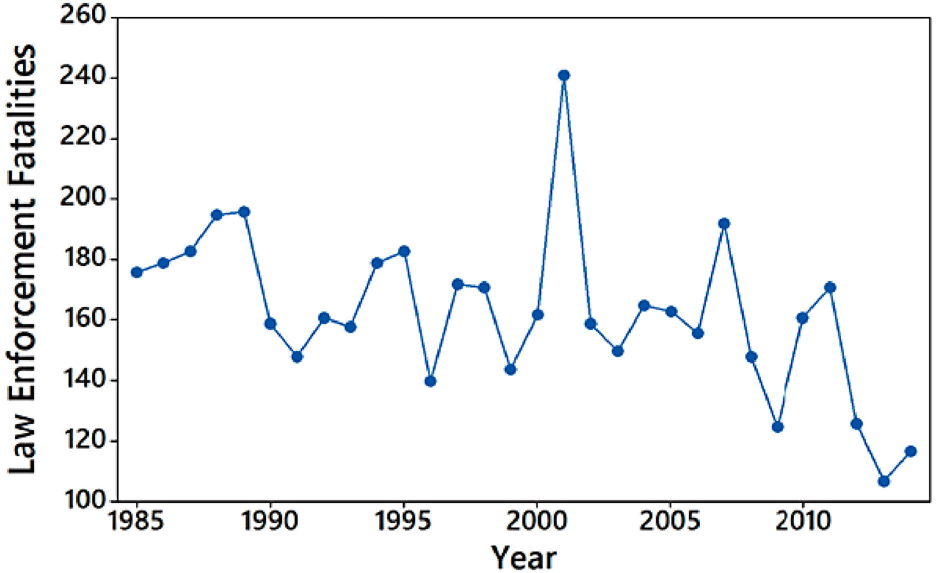
```
4  | 02                      ← Pulse rates are 40 and 42
5  | 00222224444446688888888
6  | 000000022222222222444444444444466666666666666666688888
7  | 00000000222222222222224444444446666666888888
8  | 0000002222222244444666688
9  | 02466                   ← Pulse rates are 90, 92, 94, 96, 96
10 | 24
```

Stemplots (2 of 2)

- **Stemplots (or stem-and-leaf plot)**
  - **Features of a Stemplot**
    - Shows the shape of the distribution of the data.
    - Retains the original data values.
    - The sample data are sorted (arranged in order).

Time-Series Graph (1 of 2)

- **Time-Series Graph**
  - A graph of **time-series data,** which are quantitative data that have been collected at different points in time, such as monthly or yearly



Time-Series Graph (2 of 2)

- **Time-Series Graph**
  - **Feature of a Time-Series Graph**
    - Reveals information about trends over time.

Bar Graph (1 of 2)

- **Bar Graphs**
  - A graph of bars of equal width to show frequencies of categories of **categorical** (or qualitative) data. The bars may or may not be separated by small gaps.

Bar Graph (2 of 2)

- **Bar Graphs**
  - **Feature of a Bar Graph**
    - Shows the relative distribution of categorical data so that it is easier to compare the different categories.

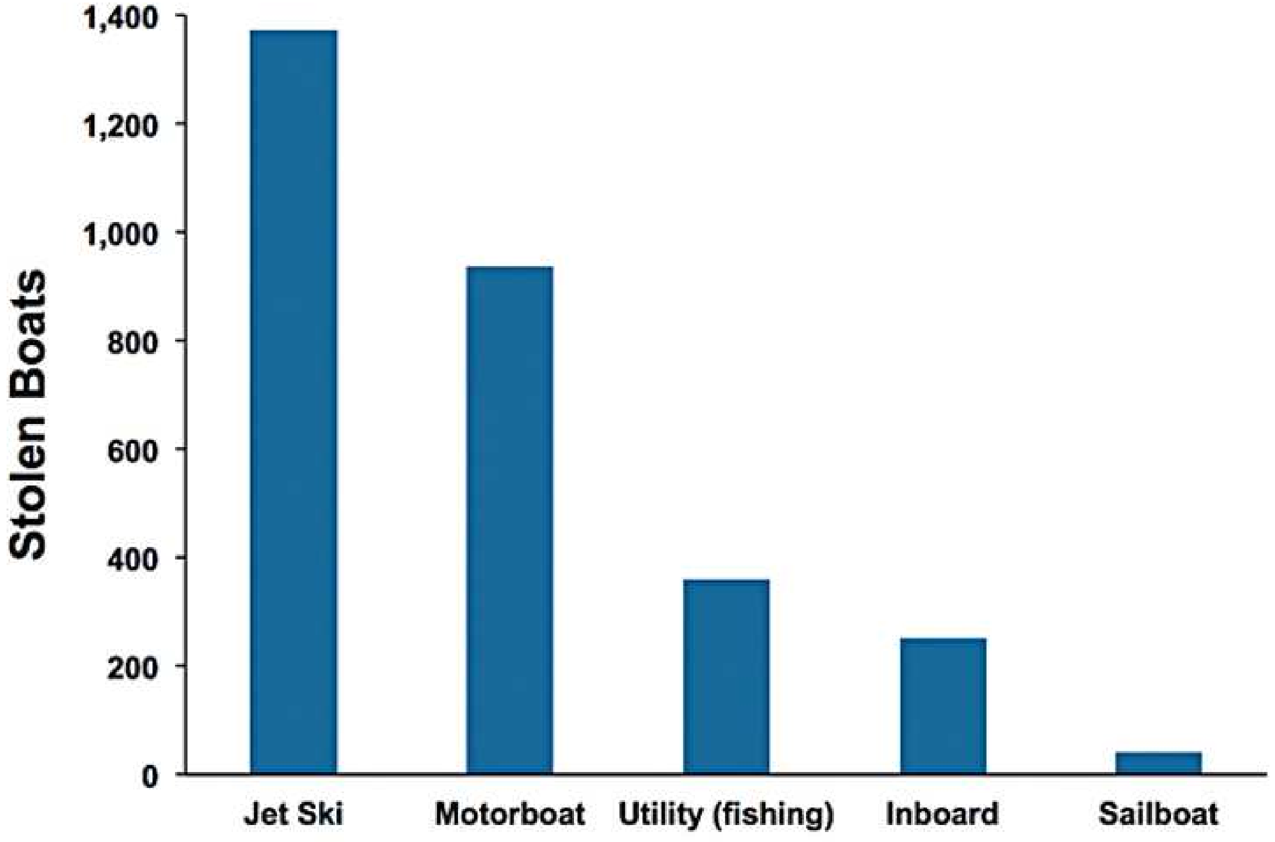
Pareto Chart (1 of 3)

- **Pareto Charts**
  - A Pareto chart is a bar graph for categorical data, with the added stipulation that the **bars are arranged in descending order** according to frequencies, so the bars decrease in height from left to right.

Pareto Chart (2 of 3)

- **Pareto Charts**
  - **Features of a Pareto Chart**
    - Shows the relative distribution of categorical data so that it is easier to compare the different categories.
    - Draws attention to the more important categories.

Pareto Chart (3 of 3)



Pareto Chart of Stolen Boats

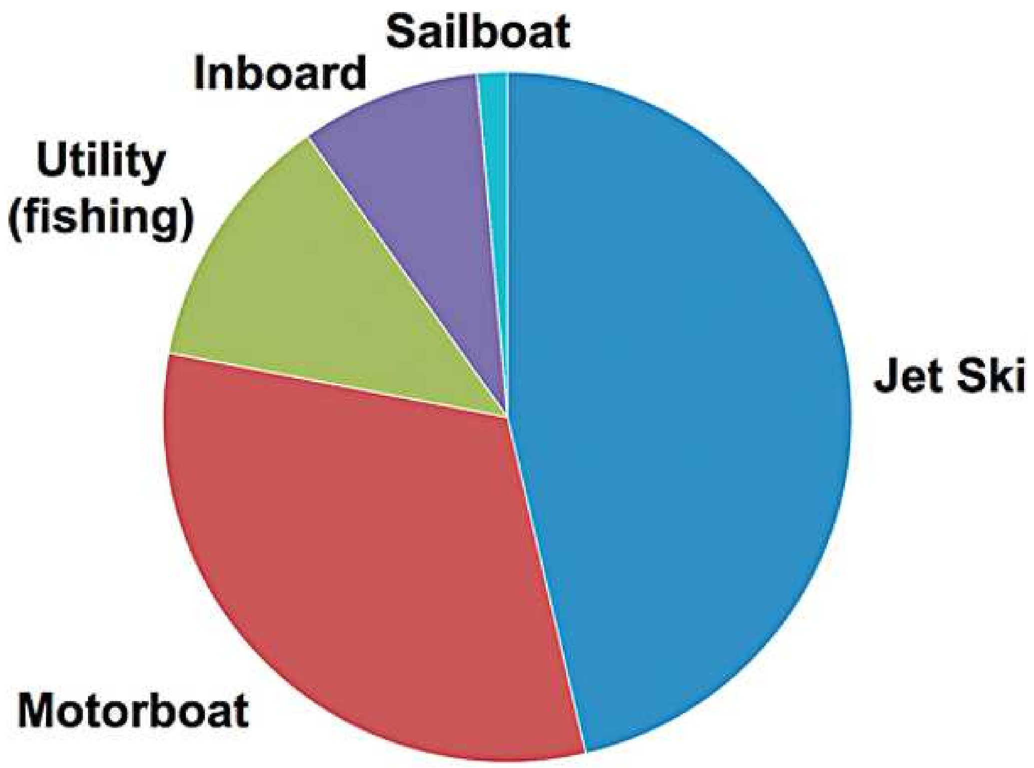
Pie Chart (1 of 3)

- **Pie Charts**
  - A very common graph that depicts categorical data as slices of a circle, in which the size of each slice is proportional to the frequency count for the category

Pie Chart (2 of 3)

- **Pie Charts**
  - **Feature of a Pie Chart**
    - Shows the distribution of categorical data in a commonly used format.

Pie Chart (3 of 3)



Pie Chart of Stolen Boats

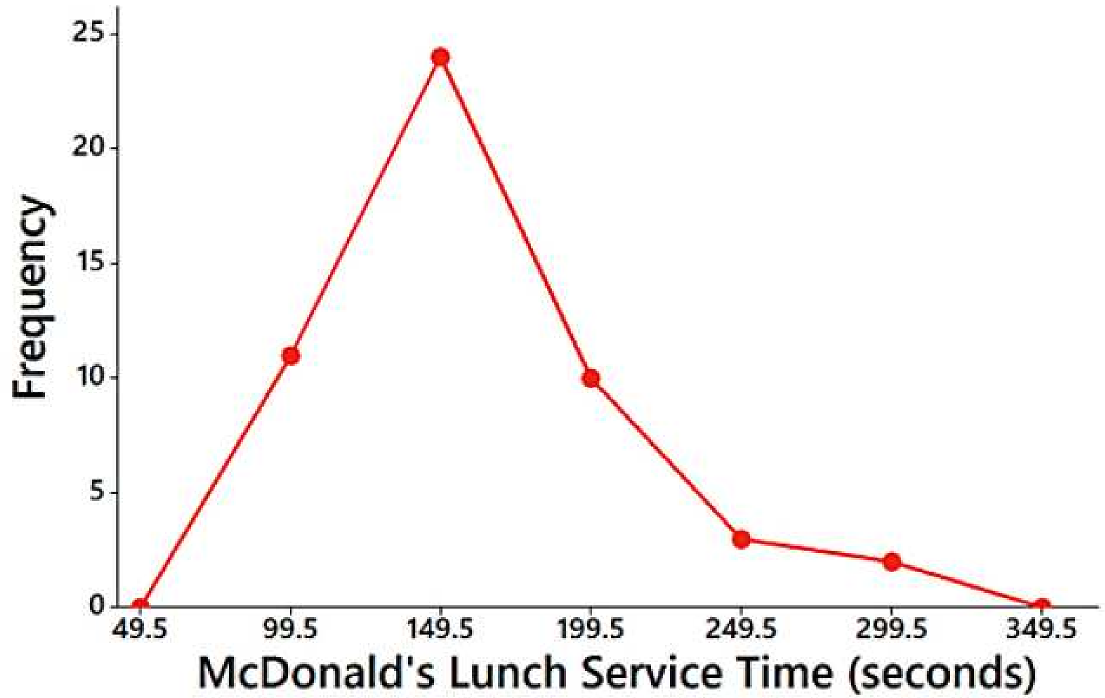
Frequency Polygon (1 of 3)

- **Frequency Polygon**
  - A graph using line segments connected to points located directly above class midpoint values
  - A frequency polygon is very similar to a histogram, but a frequency polygon uses line segments instead of bars.

Frequency Polygon (2 of 3)

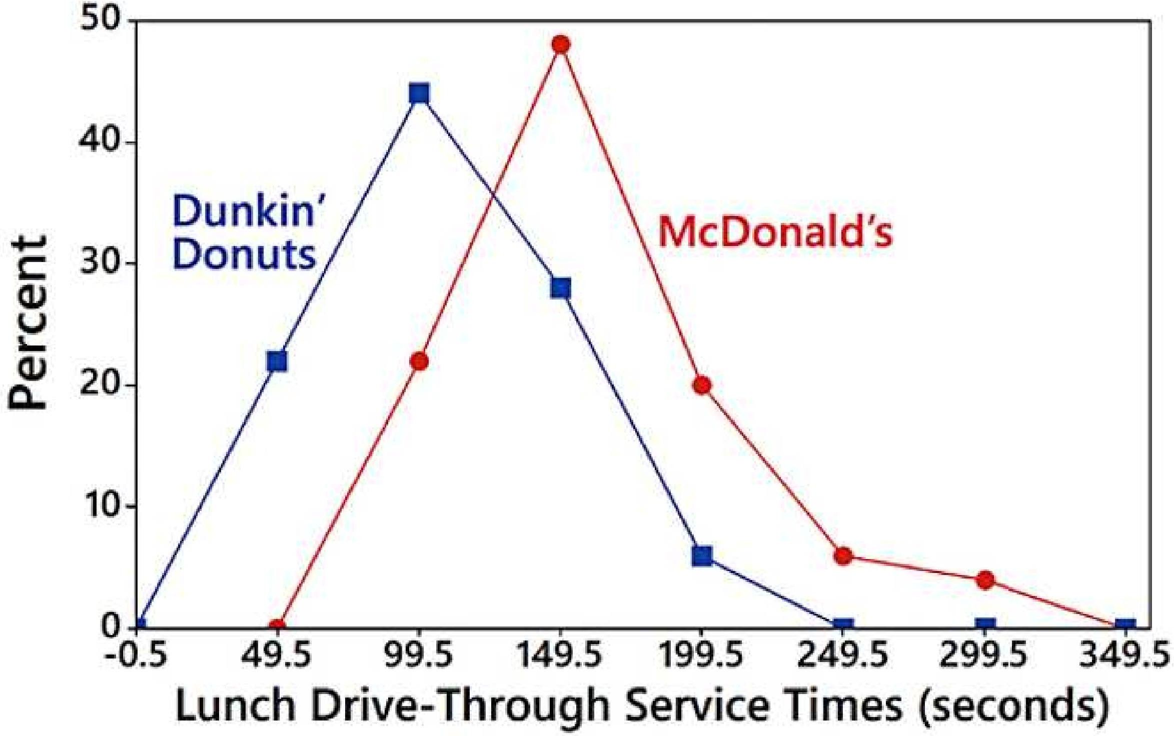
- **Frequency Polygon**
  - A variation of the basic frequency polygon is the relative frequency polygon, which uses relative frequencies (proportions or percentages) for the vertical scale.

Frequency Polygon (3 of 3)



Frequency Polygon of McDonald's Lunch Service Times

Relative Frequency Polygon



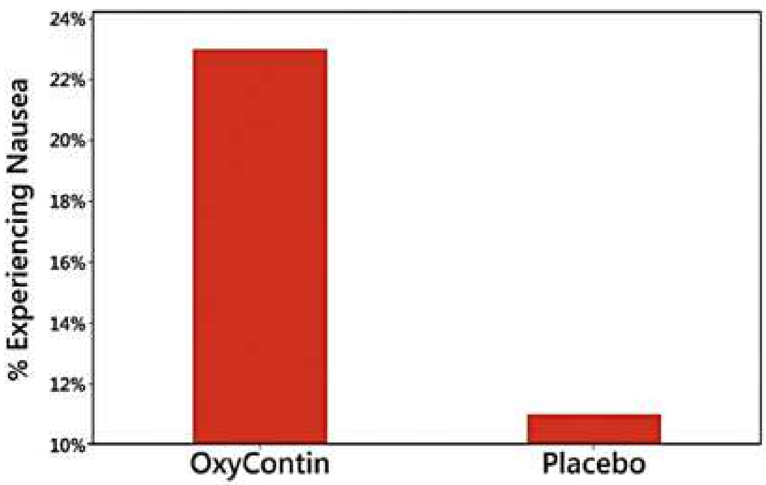
Relative Frequency Polygons for McDonald's and Dunkin' Donuts

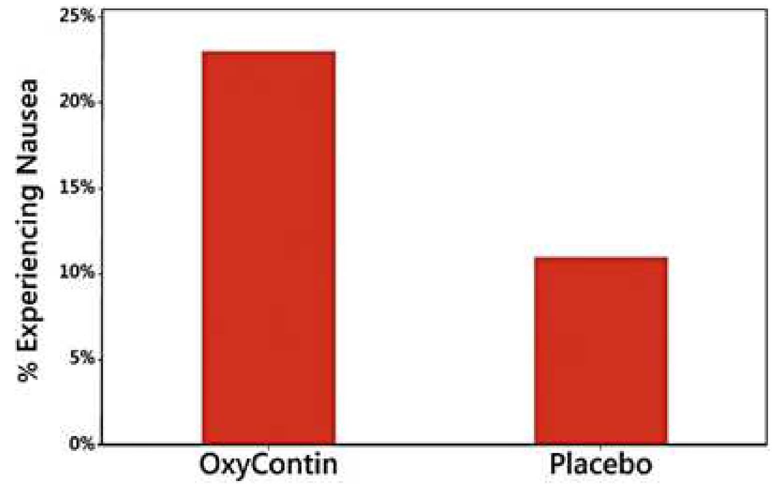
Graphs That Deceive (1 of 4)

- **Nonzero Vertical Axis**
  - A common deceptive graph involves using a vertical scale that starts at some value greater than zero to exaggerate differences between groups.

Graphs That Deceive (2 of 4)

- **Nonzero Vertical Axis**





Always examine a graph carefully to see whether a vertical axis begins at some point other than zero so that differences are exaggerated.

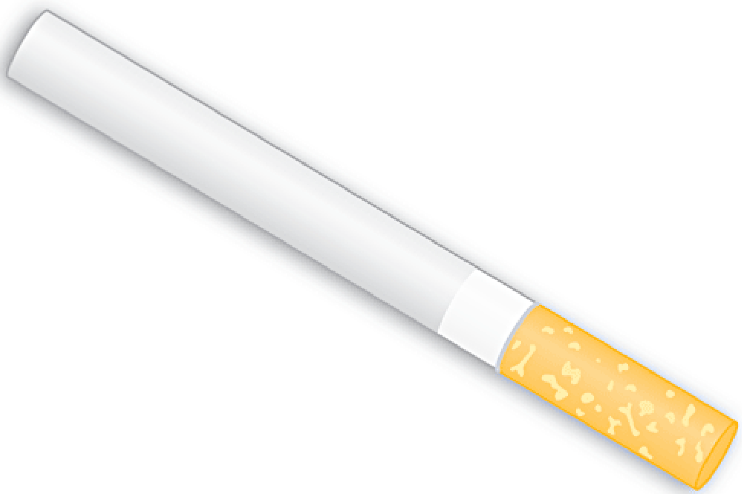
Graphs That Deceive (3 of 4)

- **Pictographs**
  - Drawings of objects, called **pictographs**, are often misleading. Data that are one-dimensional in nature (such as budget amounts) are often depicted with two-dimensional objects (such as dollar bills) or three-dimensional objects (such as stacks of coins, homes, or barrels).

Graphs That Deceive (4 of 4)

- **Pictographs**
  - By using pictographs, artists can create false impressions that grossly distort differences by using these simple principles of basic geometry:
    - When you double each side of a square, its area doesn't merely double; it increases by a factor of **four**.
    - When you double each side of a cube, its volume doesn't merely double; it increases by a factor of **eight.**

Pictographs



1970: 37% of U.S. adults smoked.

2013: 18% of U.S. adults smoked.

Concluding Thoughts (1 of 2)

In addition to the graphs we have discussed in this section, there are many other useful graphs - some of which have not yet been created. The world needs more people who can create original graphs that enlighten us about the nature of data.

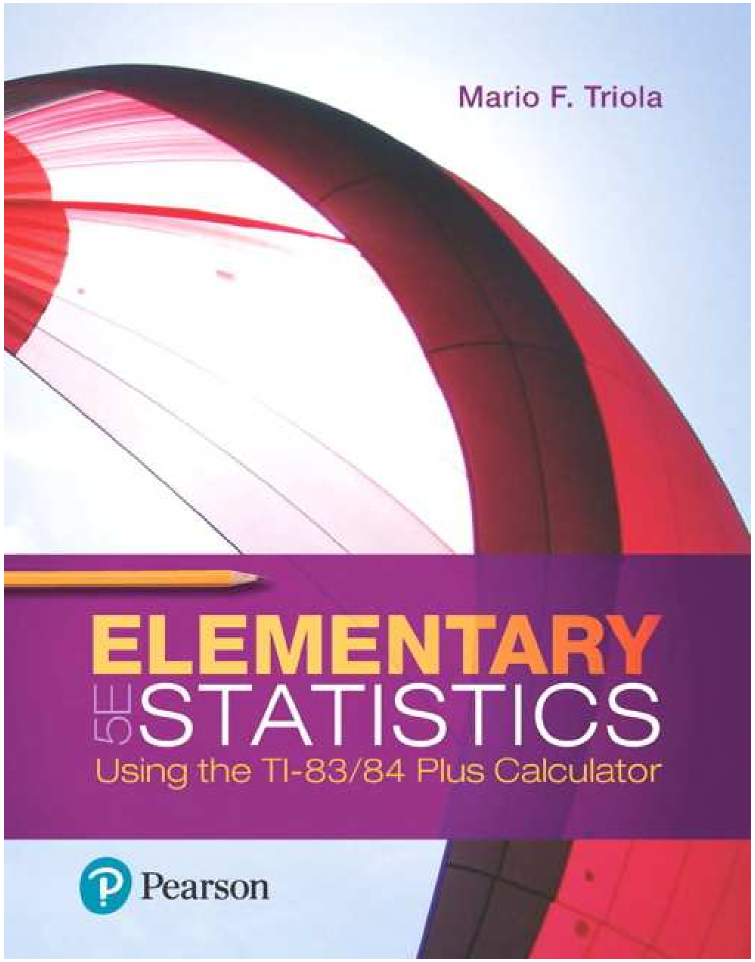
Concluding Thoughts (2 of 2)

In **The Visual Display of Quantitative Information,** Edward Tufte offers these principles:

- For small data sets of 20 values or fewer, use a table instead of a graph.
- A graph of data should make us focus on the true nature of the data, not on other elements, such as eye-catching but distracting design features.
- Do not distort data; construct a graph to reveal the true nature of the data.
- Almost all of the ink in a graph should be used for the data, not for other design elements.

Elementary Statistics Using the TI-83/84 Plus Calculator

Fifth Edition



Chapter 2

Exploring Data with Tables and Graphs

Exploring Data with Tables and Graphs

2-1 Frequency Distributions for Organizing and Summarizing Data

2-2 Histograms

2-3 Graphs that Enlighten and Graphs that Deceive

2-4 Scatterplots, Correlation, and Regression

Key Concept

Introduce the analysis of **paired** sample data.

Discuss **correlation** and the role of a graph called a **scatterplot,** and provide an introduction to the use of the **linear correlation coefficient.**

Provide a very brief discussion of **linear regression,** which involves the equation and graph of the straight line that best fits the sample paired data.

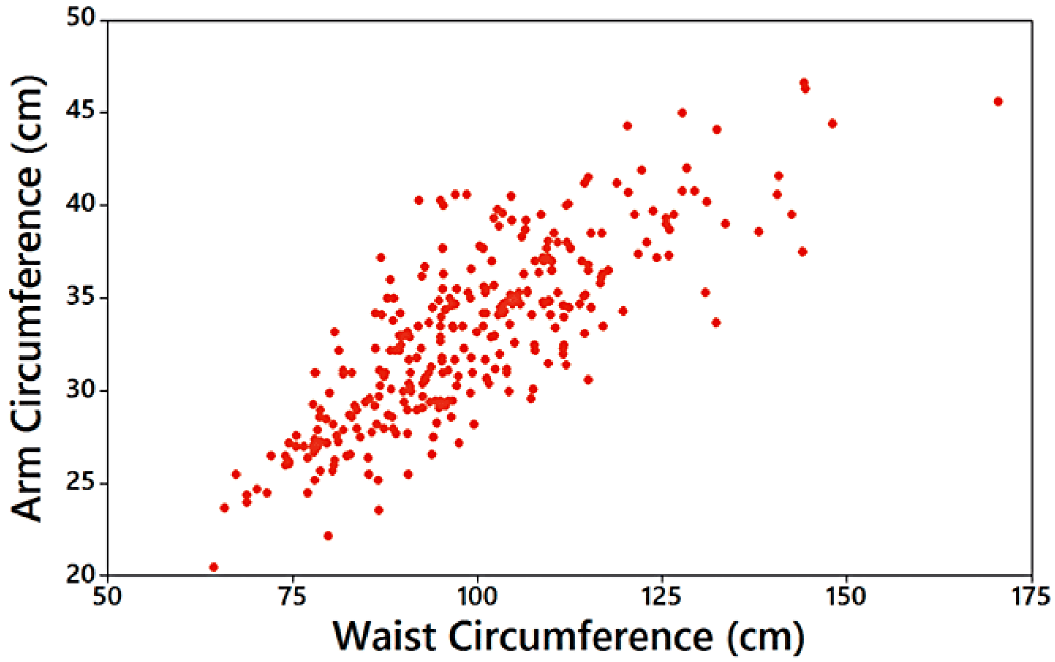
Scatterplot and Correlation (1 of 2)

- **Correlation**
  - A **correlation** exists between two variables when the values of one variable are somehow associated with the values of the other variable.
- **Linear Correlation**
  - A **linear correlation** exists between two variables when there is a correlation and the plotted points of paired data result in a pattern that can be approximated by a straight line.

Scatterplot and Correlation (2 of 2)

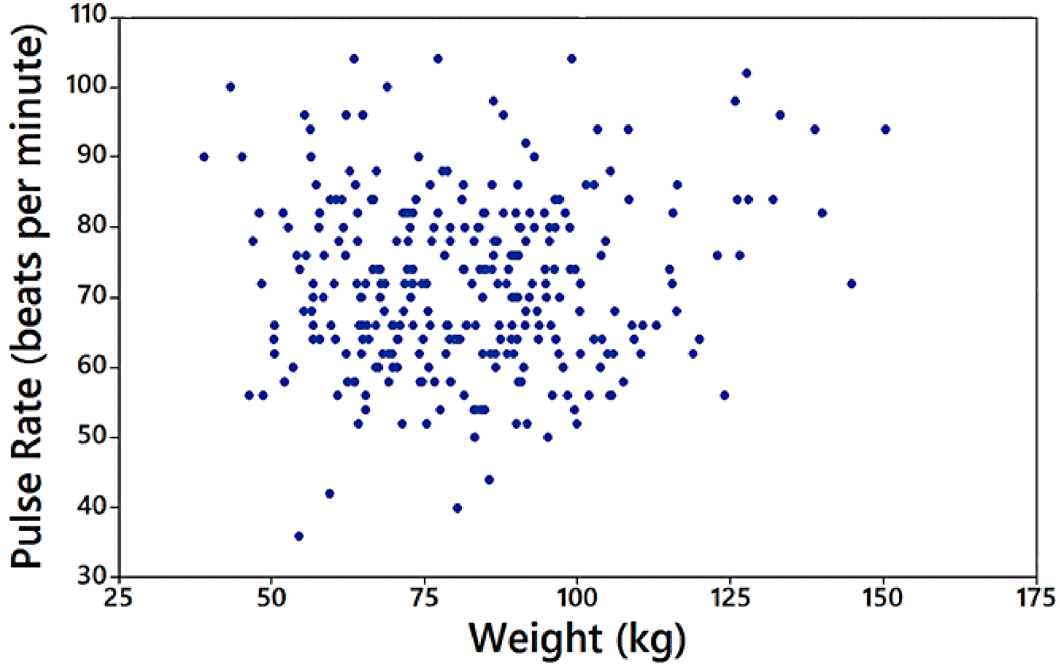
- **Scatterplot** (or **Scatter Diagram**)
  - A **scatterplot** (or **scatter diagram**) is a plot of paired (x, y) quantitative data with a horizontal x-axis and a vertical y-axis. The horizontal axis is used for the first variable (x), and the vertical axis is used for the second variable (y).

Example: Waist and Arm Correlation (1 of 2)



- **Correlation:** The distinct pattern of the plotted points suggests that there is a correlation between waist circumferences and arm circumferences.

Example: Waist and Arm Correlation (2 of 2)



- **No Correlation:** The plotted points do not show a distinct pattern, so it appears that there is no correlation between weights and pulse rates.

Linear Correlation Coefficient *r*

- **Linear Correlation Coefficient *r***
  - The **linear correlation coefficient** is denoted by r, and it measures the strength of the linear association between two variables.

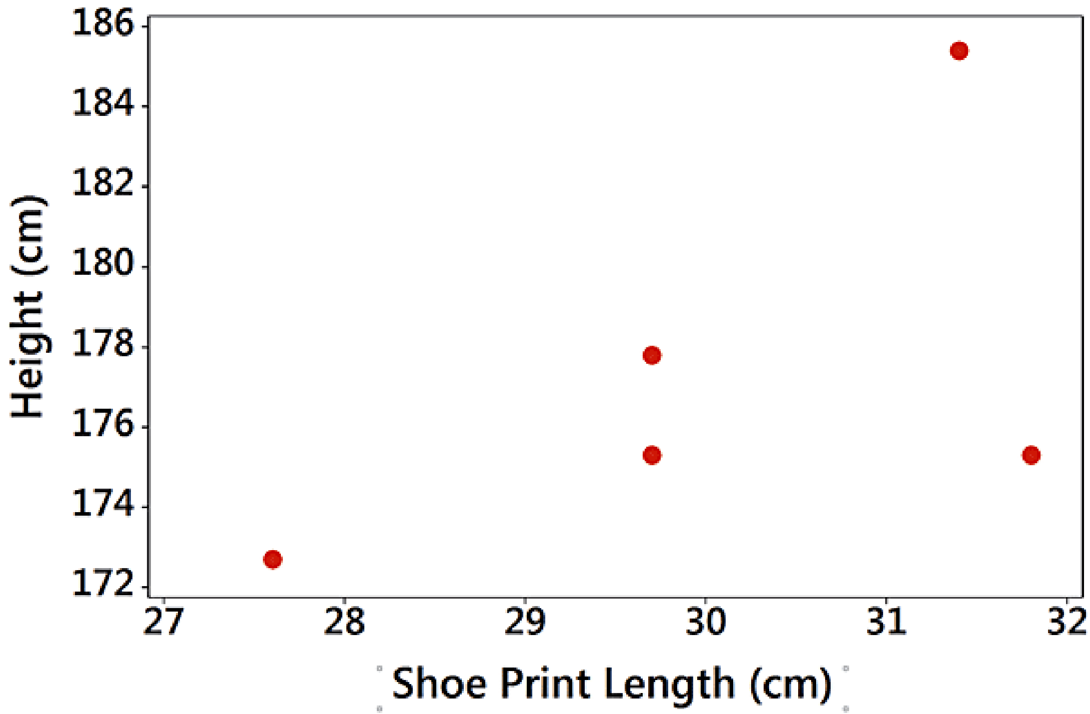
Using *r* for Determining Correlation

The computed value of the linear correlation coefficient, r, is always between −1 and 1.

- If r is close to −1 or close to 1, there appears to be a correlation.
- If r is close to 0, there does not appear to be a linear correlation.

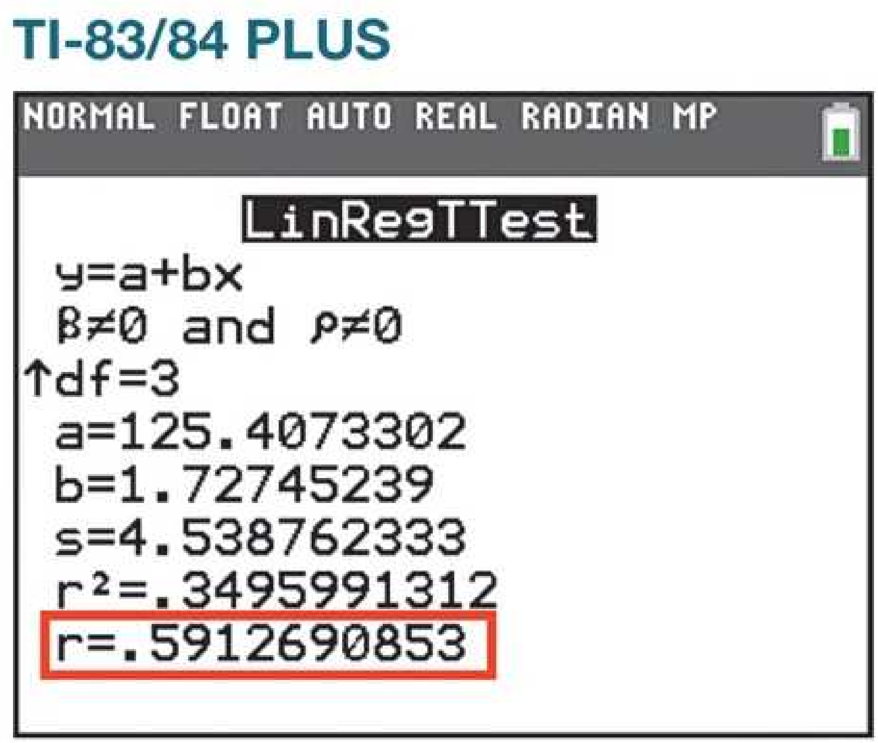
Example: Correlation between Shoe Print Lengths and Heights? (1 of 2)

Shoe Print Length (cm)	29.7	29.7	31.4	31.8	27.6
Height (cm)	175.3	177.8	185.4	175.3	172.7



Example: Correlation between Shoe Print Lengths and Heights? (2 of 2)

It isn't very clear whether there is a linear correlation.



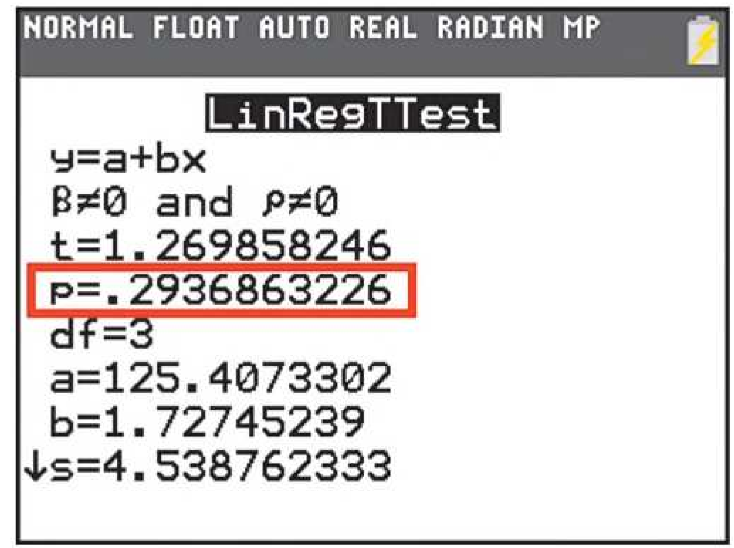
***P*-Value**

- ***P*-Value**
  - If there really is no linear correlation between two variables, the ***P*-value** is the probability of getting paired sample data with a linear correlation coefficient *r* that is at least as extreme as the one obtained from the paired sample data.

Interpreting a *P*-Value from the Previous Example

The *P*-value of 0.294 is high. It shows there is a high chance of getting a linear correlation coefficient of $r = 0.591$ (or more extreme) by chance when there is no linear correlation between the two variables.

TI-83/84 PLUS



Interpreting a P-Value from the Example Where $n = 5$

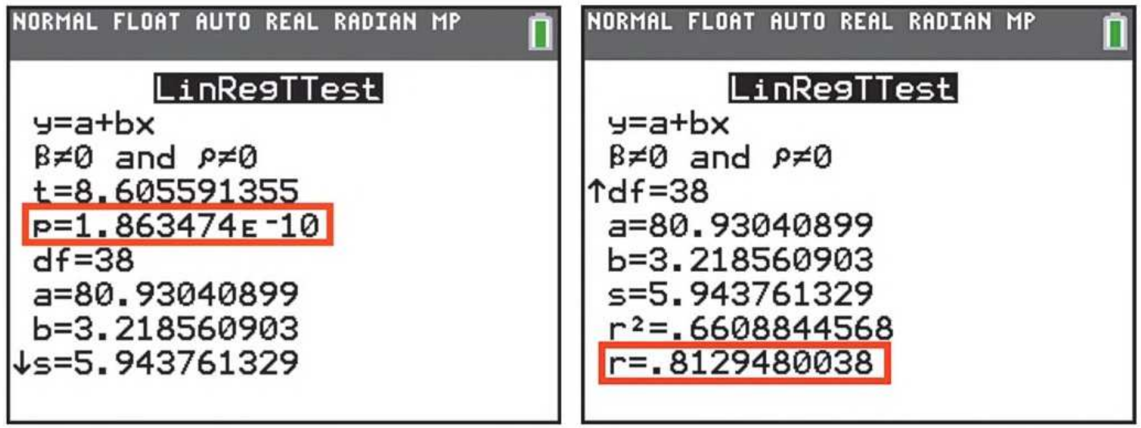
Because the likelihood of getting $r = 0.591$ or a more extreme value is so high (29.4% chance), we conclude there is not sufficient evidence to conclude there is a linear correlation between shoe print lengths and heights.

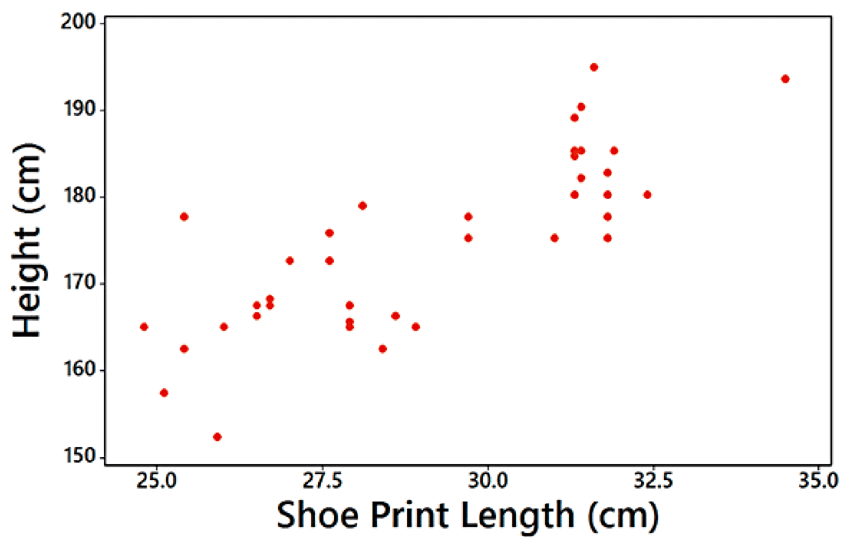
Interpreting a *P*-Value

Only a **small** *P*-value, such as 0.05 or less (or a 5% chance or less), suggests that the sample results are **not** likely to occur by chance when there is no linear correlation, so a small *P*-value supports a conclusion that there is a linear correlation between the two variables.

Example: Correlation between Shoe Print Lengths and Heights (n = 40)

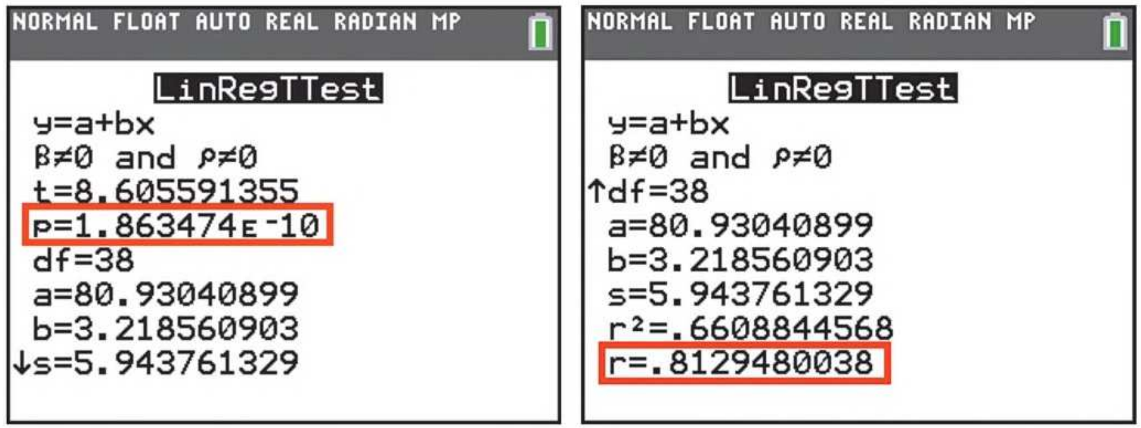
TI-83/84 PLUS





Example: Correlation between Shoe Print Lengths and Heights

TI-83/84 PLUS



The scatterplot shows a distinct pattern. The value of the linear correlation coefficient is $r = 0.813$, and the *P*-value is 0.000. Because the *P*-value of 0.000 is **small**, we have sufficient evidence to conclude there is a linear correlation between shoe print lengths and heights.

Regression

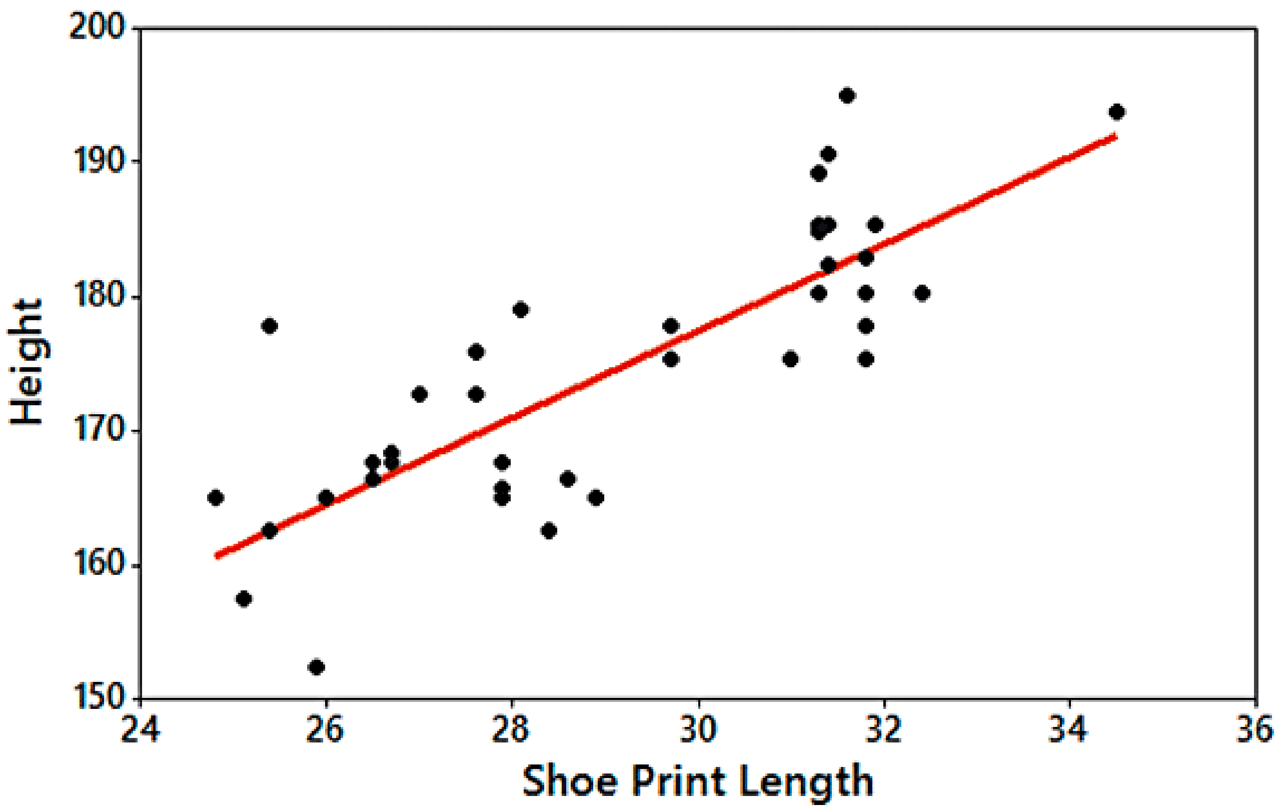
- **Regression**
  - Given a collection of paired sample data, the **regression line** (or **line of best fit,** or **least-squares line**) is the straight line that "best" fits the scatterplot of the data.

The **regression equation**

$$\hat{y} = b_0 + b_1 x$$

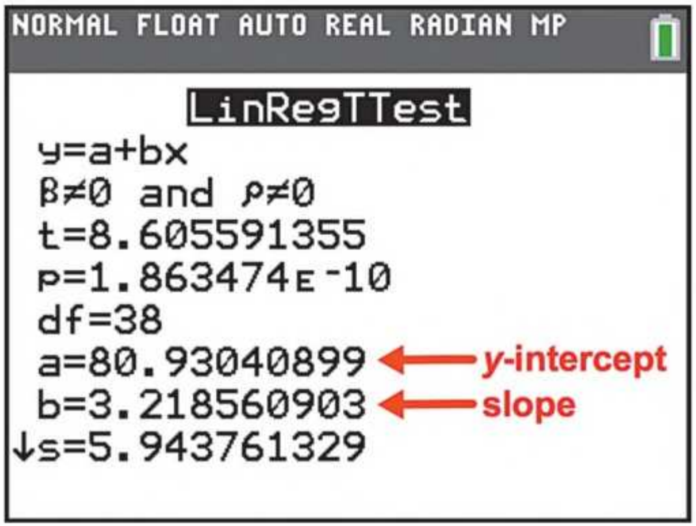
algebraically describes the regression line.

Example: Regression Line (1 of 2)



Example: Regression Line (2 of 2)

TI-83/84 PLUS



The general form of the regression equation has a y-intercept of $b_0 = 80.9$ and slope $b_1 = 3.22$.

The equation of the regression line is $\hat{y} = 80.9 + 3.22x$.

Using variable names, the equation is:

$$\text{Height} = 80.9 + 3.22\ (\text{Shoe Print Length})$$