Chapter 2, Part A
Descriptive Statistics: Tabular and Graphical Displays

- Summarizing Data for a Categorical Variable
  - Categorical data use labels or names to identify categories of like items.
- Summarizing Data for a Quantitative Variable
  - Quantitative data are numerical values that indicate how much or how many.

Summarizing Data for a Categorical Variable

- Frequency Distribution
- Relative Frequency Distribution
- Percent Frequency Distribution
- Bar Chart
- Pie Chart

Frequency Distribution (1 of 3)

- A frequency distribution is a tabular summary of data showing the number (frequency) of observations in each of several non-overlapping categories or classes.
- The objective is to provide insights about the data that cannot be quickly obtained by looking only at the original data.

Frequency Distribution (2 of 3)

Example

- Soft drink purchasers were asked to select one among the five popular soft drinks: Coca-Cola, Diet Coke, Dr. Pepper, Pepsi, and Sprite.
- Soft drinks selected by a sample of 20 purchasers are:

Coca-Cola	Pepsi	Dr. Pepper
Diet Coke	Dr. Pepper	Dr. Pepper
Dr. Pepper	Pepsi	Pepsi
Pepsi	Coca-Cola	Diet Coke
Pepsi	Diet Coke	Dr. Pepper
Pepsi	Pepsi	Sprite
Pepsi	Pepsi	

Frequency Distribution (3 of 3)

Example

Rating	Frequency
Coca-Cola	2
Diet Coke	3
Dr. Pepper	5
Pepsi	9
Sprite	1
Total	20

Using Excel's COUNTIF Function to Construct a Frequency Distribution (1 of 2)

- Excel Formula worksheet

	A	B	C	D
1	**Soft drink selected**		**Soft drink**	**Frequency**
2	Coca-Cola		Coca-Cola	= COUNTIF(A$2:A$21,C2)
3	Diet coke		Diet coke	= COUNTIF(A$2:A$21,C3)
4	Dr. Pepper		Dr. Pepper	= COUNTIF(A$2:A$21,C4)
5	Pepsi		Pepsi	= COUNTIF(A$2:A$21,C5)
6	Pepsi		Sprite	= COUNTIF(A$2:A$21,C6)
7	Pepsi		Total	= SUM(D2:D6)
8	Pepsi			

Note: Rows 9 to 21 are not shown.

Using Excel's COUNTIF Function to Construct a Frequency Distribution (2 of 2)

- Excel Value Worksheet

	A	B	C	D
1	**Soft drink selected**		**Soft drink**	**Frequency**
2	Coca-Cola		Coca-Cola	2
3	Diet coke		Diet coke	3
4	Dr. Pepper		Dr. Pepper	5
5	Pepsi		Pepsi	9
6	Pepsi		Sprite	1
7	Pepsi		Total	20
8	Pepsi			

Note: Rows 9 to 21 are not shown.

Relative Frequency Distribution

- The relative frequency of a class is the fraction or proportion of the total number of data items belonging to a class.

$$\text{Relative frequency of a class} = \frac{\text{Frequency of the class}}{n}$$

- A relative frequency distribution is a tabular summary of a set of data showing the relative frequency for each class.

Percent Frequency Distribution

- The percent frequency of a class is the relative frequency multiplied by 100.
- A percent frequency distribution is a tabular summary of a set of data showing the percent frequency for each class.

Relative Frequency and Percent Frequency Distributions

Example

Rating	Relative Frequency	Percent Frequency
Coca-Cola	.38	38
Diet Coke	.16	16
Dr. Pepper	.10	10
Pepsi	.26	26
Sprite	.10	10
Total	1.00	100

.38(100) = 38

5/50 = 0.10

Using Excel to Construct Relative Frequency and Percent Frequency Distributions

- Excel Formula and Value Worksheets

	A	B	C	D	E
1					
2					
3	**Soft Drink**	**Frequency**	**Relative Frequency**	**Percent Frequency**	
4	Coca-Cola	19	=B4/$B$9	=C4*100	
5	Diet Coke	8	=B5/$B$9	=C5*100	
6	Dr. Pepper	5	=B6/$B$9	=C6*100	
7	Pepsi	13	=B7/$B$9	=C7*100	
8	Sprite	5	=B8/$B$9	=C8*100	
9	**Total**	**50**	=SUM(C4:C8)	=SUM(D4:D8)	
10					
11					

	A	B	C	D	E
1					
2					
3	**Soft Drink**	**Frequency**	**Relative Frequency**	**Percent Frequency**	
4	Coca-Cola	19	0.38	38	
5	Diet Coke	8	0.16	16	
6	Dr. Pepper	5	0.1	10	
7	Pepsi	13	0.26	26	
8	Sprite	5	0.1	10	
9	**Total**	**50**	1	100	
10					

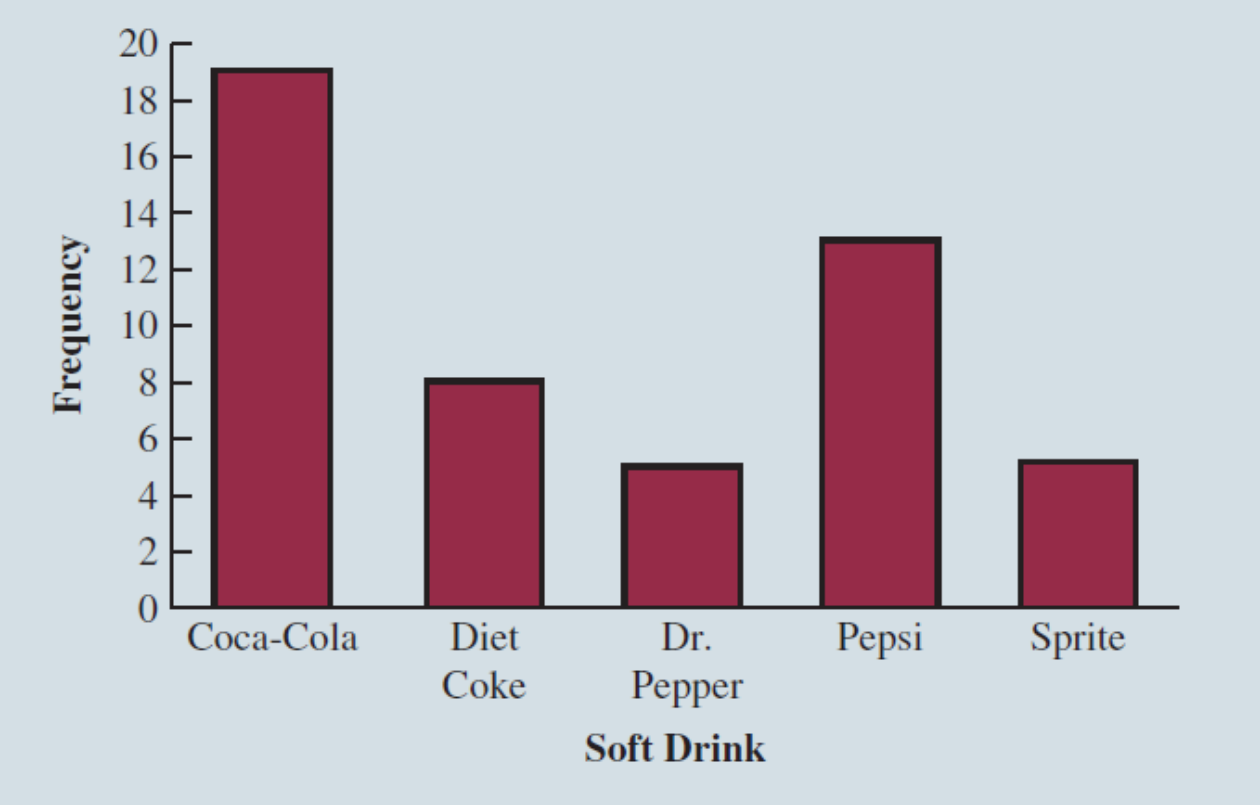
Note: Rows 9 to 21 are not shown.

Bar Chart (1 of 2)

- A bar chart is a graphical display for depicting categorical data.
- On one axis (usually the horizontal axis), we specify the labels that are used for each of the classes.
- A frequency, relative frequency, or percent frequency scale can be used for the other axis (usually the vertical axis).
- Using a bar of fixed width drawn above each class label, we extend the height appropriately.
- The bars are separated to emphasize the fact that each class is separate.

Bar Chart (2 of 2)

Bar chart of soft drink purchases



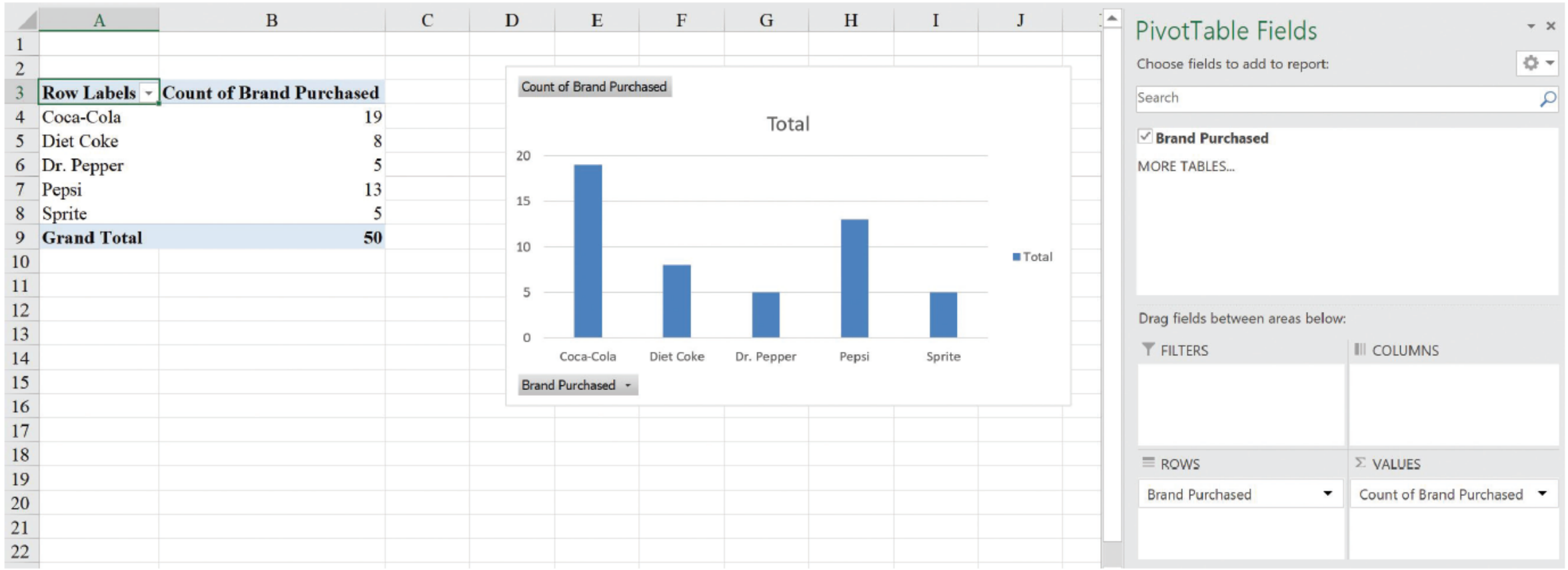
Using Excel's Recommended Charts Tool to Construct a Bar Chart (1 of 2)

- **Step 1:** Select cells C1:D6.
- **Step 2:** Click **Insert** on the Ribbon.
- **Step 3:** In the **Charts** group, click **Recommended Charts** (a preview showing of bar chart appears).
- **Step 4:** Click **OK** (the bar chart will appear in a new worksheet).

Using Excel's Recommended Charts Tool to Construct a Bar Chart (2 of 2)

- **Step 1:** Click the **Chart Title** and replace it with **Bar chart of Soft Drink Purchases.**
- **Step 2:** Click the **Chart Elements** button. When the list of chart elements appears, click **Axis Titles.**
- **Step 3:** Click the **Horizontal (Category) Axis Title** and replace it with **Soft Drink.**
- **Step 4:** Click the Vertical (Value) Axis Title and replace it with **Frequency.**

Bar Chart of Soft Drink Purchases Constructed Using Excel's Recommended Charts Tool

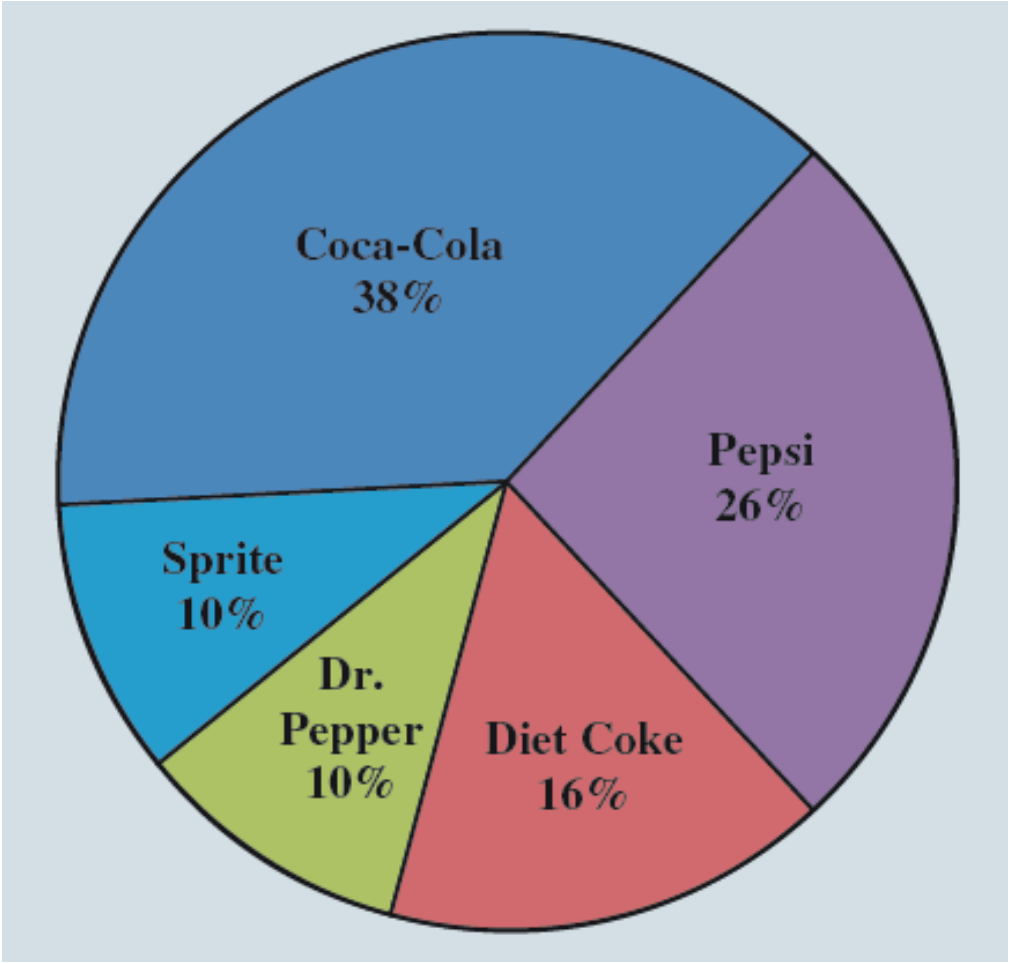


Row Labels	Count of Brand Purchased
Coca-Cola	19
Diet Coke	8
Dr. Pepper	5
Pepsi	13
Sprite	5
Grand Total	**50**

Pie Chart (1 of 2)

- The pie chart is a commonly used graphical display for presenting relative frequency and percent frequency distributions for categorical data.
- First draw a circle; then use the relative frequencies to subdivide the circle into sectors that correspond to the relative frequency for each class.
- Since there are 360 degrees in a circle, a class with a relative frequency of .25 would consume $.25(360) = 90$ degrees of the circle.

Pie Chart of Soft Drink Purchases



Pie Chart (2 of 2)

Example

Inferences from the pie chart:

- Over one-third of the customers surveyed prefer Coca-Cola.
- The second preference is for Pepsi with 26% of the customers opting for it.
- Only 10% of the customers opt for Sprite or Dr. Pepper.

Summarizing Data for a Quantitative Variable

- Frequency Distribution
- Relative Frequency and Percent Frequency Distributions
- Dot Plot
- Histogram
- Cumulative Distributions
- Stem-and-Leaf Display

Frequency Distribution (1 of 9)

Example

Sanderson and Clifford, a small public accounting firm, wants to determine time in days required to complete year-end audits. It takes a sample of 20 clients.

Frequency Distribution (2 of 9)

Example: Sanderson and Clifford

Year-End Audit Time (in Days)

12	14	19	18
15	15	18	17
20	27	22	23
22	21	33	28
14	18	16	13

Frequency Distribution (3 of 9)

The three steps necessary to define the classes for a frequency distribution with quantitative data are:

- **Step 1:** Determine the number of non-overlapping classes.
- **Step 2:** Determine the width of each class.
- **Step 3:** Determine the class limits.

Frequency Distribution (4 of 9)

Guidelines for Determining the Number of Classes

- Use between 5 and 20 classes.
- Data sets with a larger number of elements usually require a larger number of classes.
- Smaller data sets usually require fewer classes.
- The goal is to use enough classes to show the variation in the data, but not so many classes that some contain only a few data items.

Frequency Distribution (5 of 9)

Guidelines for Determining the Width of Each Class

- Use classes of equal width.

$$\text{Approximate Class Width} = \frac{\text{Largest data value} - \text{Smallest data value}}{\text{Number of classes}}$$

- Making the classes the same width reduces the chance of inappropriate interpretations.

Frequency Distribution (6 of 9)

Note on Number of Classes and Class Width

- In practice, the number of classes and the appropriate class width are determined by trial and error.
- Once a possible number of classes is chosen, the appropriate class width is found.
- The process can be repeated for a different number of classes.
- Ultimately, the analyst uses judgment to determine the combination of the number of classes and class width that provides the best frequency distribution for summarizing the data.

Frequency Distribution (7 of 9)

Guidelines for Determining the Class Limits

- Class limits must be chosen so that each data item belongs to one and only one class.
- The lower class limit identifies the smallest possible data value assigned to the class.
- The upper class limit identifies the largest possible data value assigned to the class.
- The appropriate values for the class limits depend on the level of accuracy of the data.
- An open-end class requires only a lower class limit or an upper class limit.

Frequency Distribution (8 of 9)

Class Midpoint

- In some cases, we want to know the midpoints of the classes in a frequency distribution for quantitative data.
- The class midpoint is the value halfway between the lower and upper class limits.

Frequency Distribution (9 of 9)

Example: Sanderson and Clifford

- If we choose five classes:

$$\text{Approximate Class Width} = \frac{(33-12)}{5} = 4.2 \cong 4$$

Time in days	Frequency
10 to 14	4
15 to 19	8
20 to 24	5
25 to 29	2
30 to 34	1
Total	20

Using Excel's PivotTable to Construct a Frequency Distribution (1 of 4)

- **Step 1:** Select any cell in the data set.
- **Step 2:** Click **Insert** on the Ribbon.
- **Step 3:** In the **Tables** group click **Pivot Table.**
- **Step 4:** When the **Create PivotTable** dialog box appears, click **OK** (a PivotTable and PivotTable Fields dialog box will appear in a new worksheet).

Using Excel's PivotTable to Construct a Frequency Distribution (2 of 4)

- **Step 5:** In the **PivotTable Fields** dialog box:

  Drag **Audit time** to the **ROWS** area.

  Drag **Audit time** to the **VALUES** area.

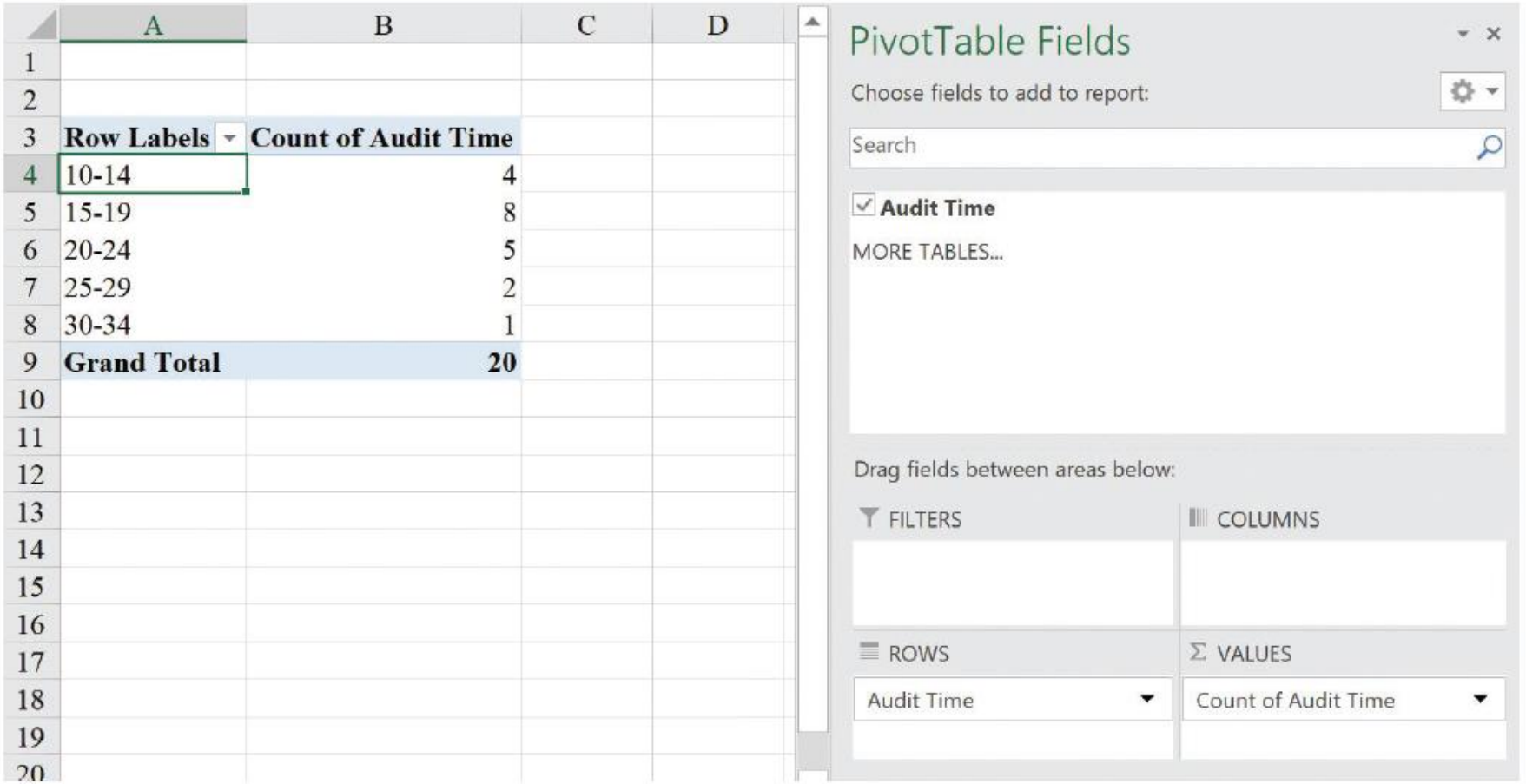
- **Step 6:** Click on **Sum of Audit time** in the **VALUES** area.
- **Step 7:** Click **Value Field Settings…** from the list of options.
- **Step 8:** When the **Value Field Settings** dialog box appears, under **Summarize value field by,** choose **Count** and click **OK**.

Using Excel's PivotTable to Construct a Frequency Distribution (3 of 4)

To construct the frequency distribution, we must group the rows containing audit time.

- **Step 1:** Right click any cell in the PivotTable report containing an audit time.
- **Step 2:** Choose **Group** from the list of options that appears.
- **Step 3:** When the Grouping dialog box appears:
  - Enter 10 in the **Starting at** box.
  - Enter 34 in the **Ending at** box.
  - Enter 5 in the **By** box.
  - Click **OK**.

Using Excel's PivotTable to Construct a Frequency Distribution (4 of 4)



Row Labels	Count of Audit Time
10-14	4
15-19	8
20-24	5
25-29	2
30-34	1
Grand Total	**20**

Relative Frequency and Percent Frequency Distributions (1 of 2)

Example: Sanderson and Clifford

Audit time (in days)	Relative Frequency	Percent Frequency
10 to 14	.20 (4/20)	20 (0.2 × 100)
15 to 19	.40	40
20 to 25	.25	25
25 to 29	.10	10
30 to 34	.05	5
	Total 1.00	100

Relative Frequency and Percent Frequency Distributions (2 of 2)

Example: Sanderson and Clifford

Insights obtained from the Percent Frequency Distribution:

- 40% of the audits required from 15 to 19 days.
- Another 25% of the audits required 20 to 25 days.
- Only 5% of the audits required more than 30 days.

Dot Plot (1 of 2)

- One of the simplest graphical summaries of data is a dot plot.
- A horizontal axis shows the range of data values.
- Then each data value is represented by a dot placed above the axis.

Dot Plot (2 of 2)

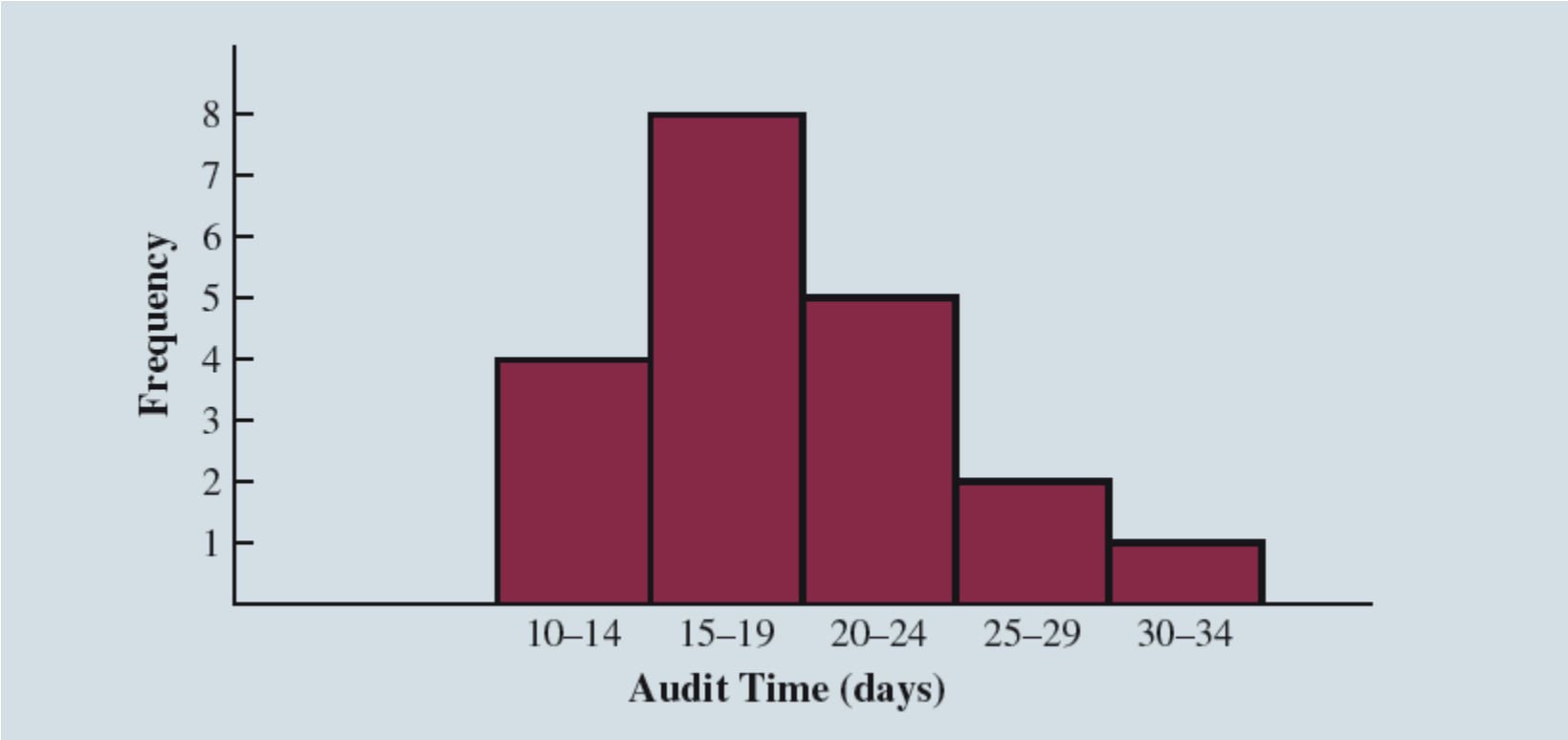
Example: Sanderson and Clifford

Histogram (1 of 2)

- Another common graphical display of quantitative data is a histogram.
- The variable of interest is placed on the horizontal axis.
- A rectangle is drawn above each class interval with its height corresponding to the interval's frequency, relative frequency, or percent frequency.
- Unlike a bar graph, a histogram has no natural separation between rectangles of adjacent classes.

Histogram (2 of 2)

Example: Sanderson and Clifford



Using Excel's Recommended Charts Tool to Construct a Histogram (1 of 4)

- **Step 1:** Select any cell in the PivotTable report.
- **Step 2:** Click the **Insert** tab on the Ribbon.
- **Step 3:** In the **Charts** group, click **Recommended Charts**.
- **Step 4:** Click **OK.**

Using Excel's Recommended Charts Tool to Construct a Histogram (2 of 4)

- **Step 1:** Right-click any bar in the chart and choose **Format Data Series…** from the list of options.
- **Step 2:** When the **Format Data Series** task pane appears:

  Go to the **Series Options** section.

  Set the **Gap Width** to 0.

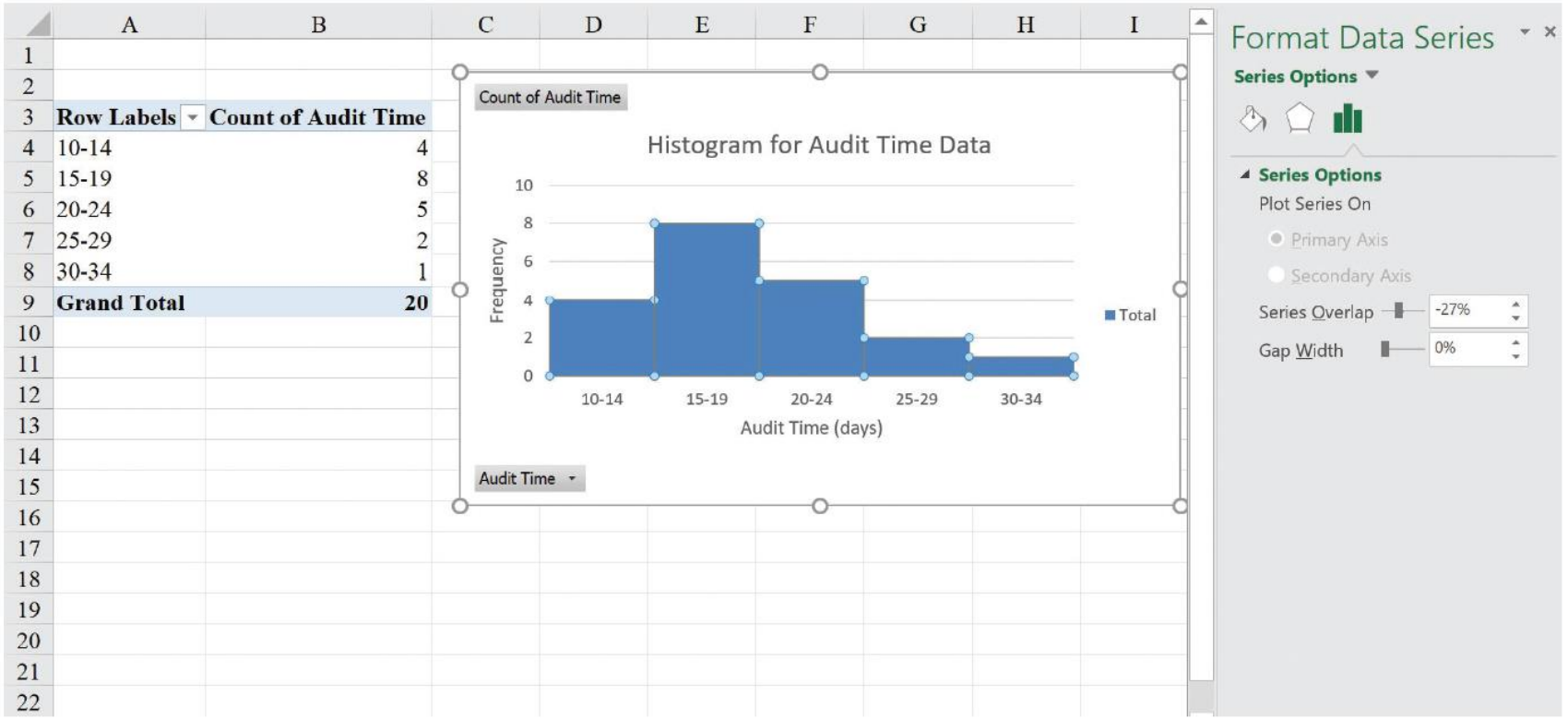
  Click the **Close** button at the top right of the task pane.
- **Step 3:** Click the **Chart Title** and replace it with *Histogram for Audit Time Data*.
- **Step 4:** Click the **Chart Elements** button (in top right **corner of the chart**).

Using Excel's Recommended Charts Tool to Construct a Histogram (3 of 4)

- **Step 5:** When the list of chart elements appears, select the check box for **Axis Titles** and deselect the check box for **Legend.**
- **Step 6:** Click the **Horizontal (Category) Axis Title** and replace it with *Audit Time (days)*. Click the **Close** button at the top right.
- **Step 7:** Click the **Vertical (Value) Axis Title** and replace it with *Frequency*.

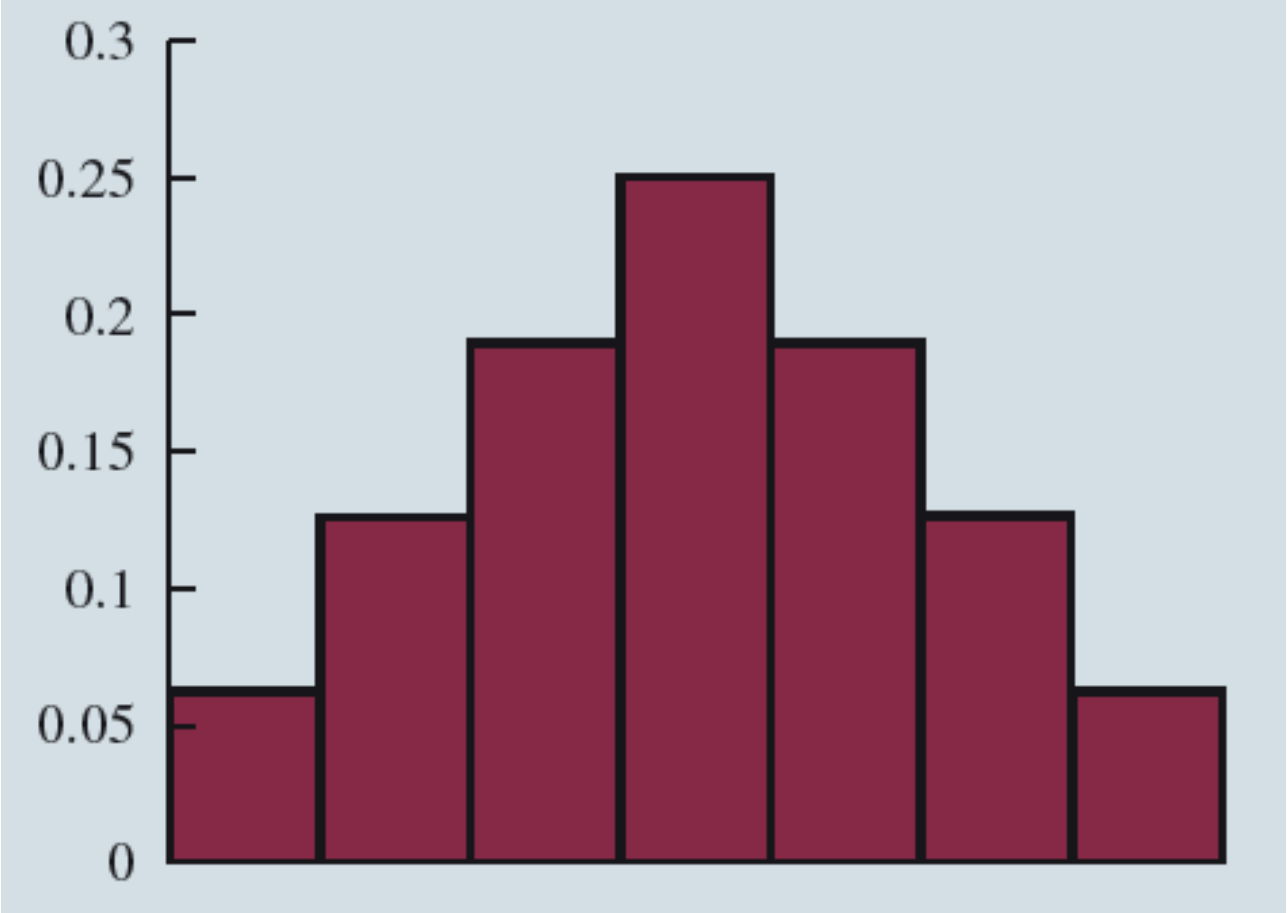
Using Excel's Recommended Charts Tool to Construct a Histogram (4 of 4)

Example



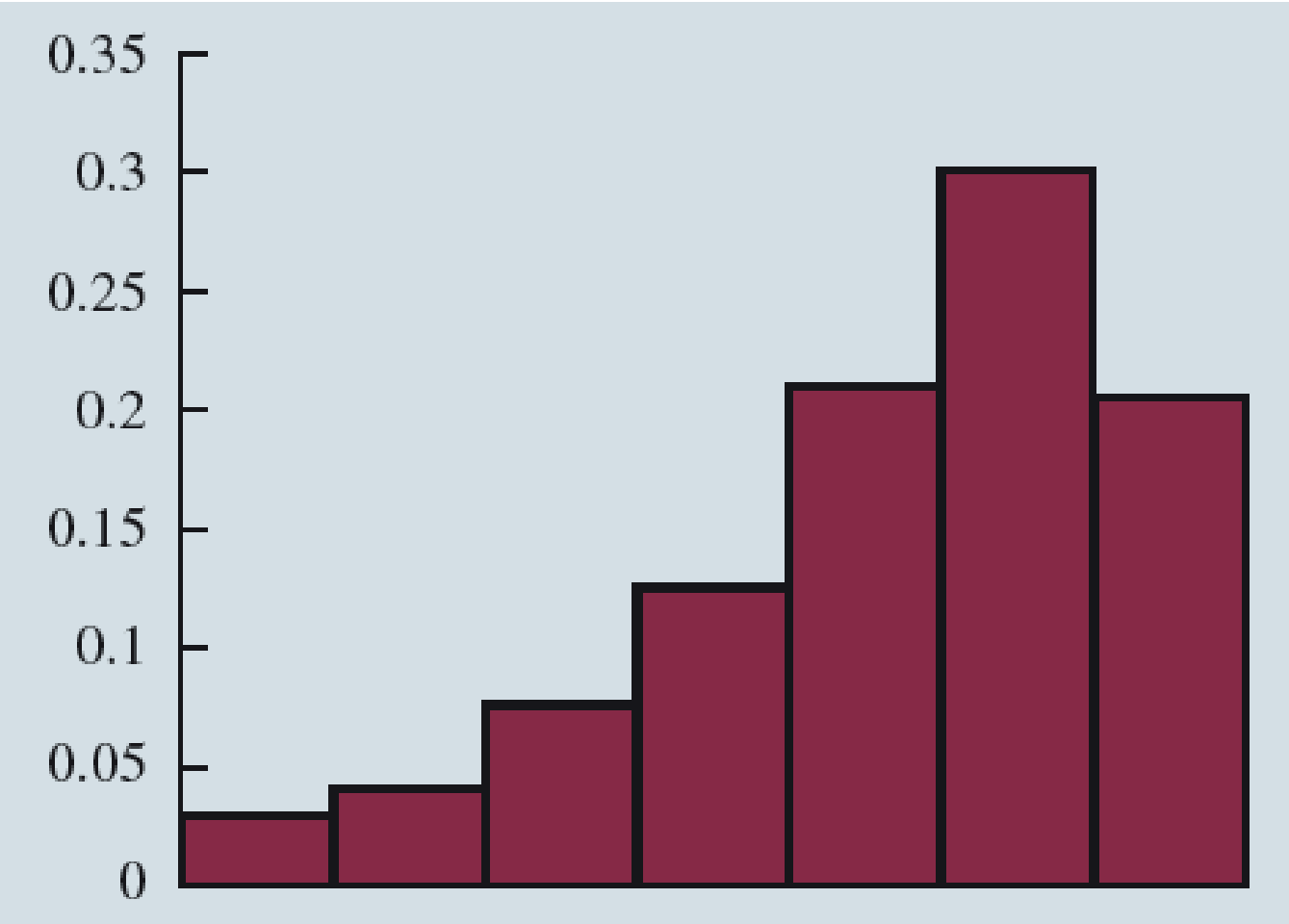
Histograms Showing Skewness (1 of 4)

- Symmetric
  - Left tail is the mirror image of the right tail.
  - **Example:** Heights of People



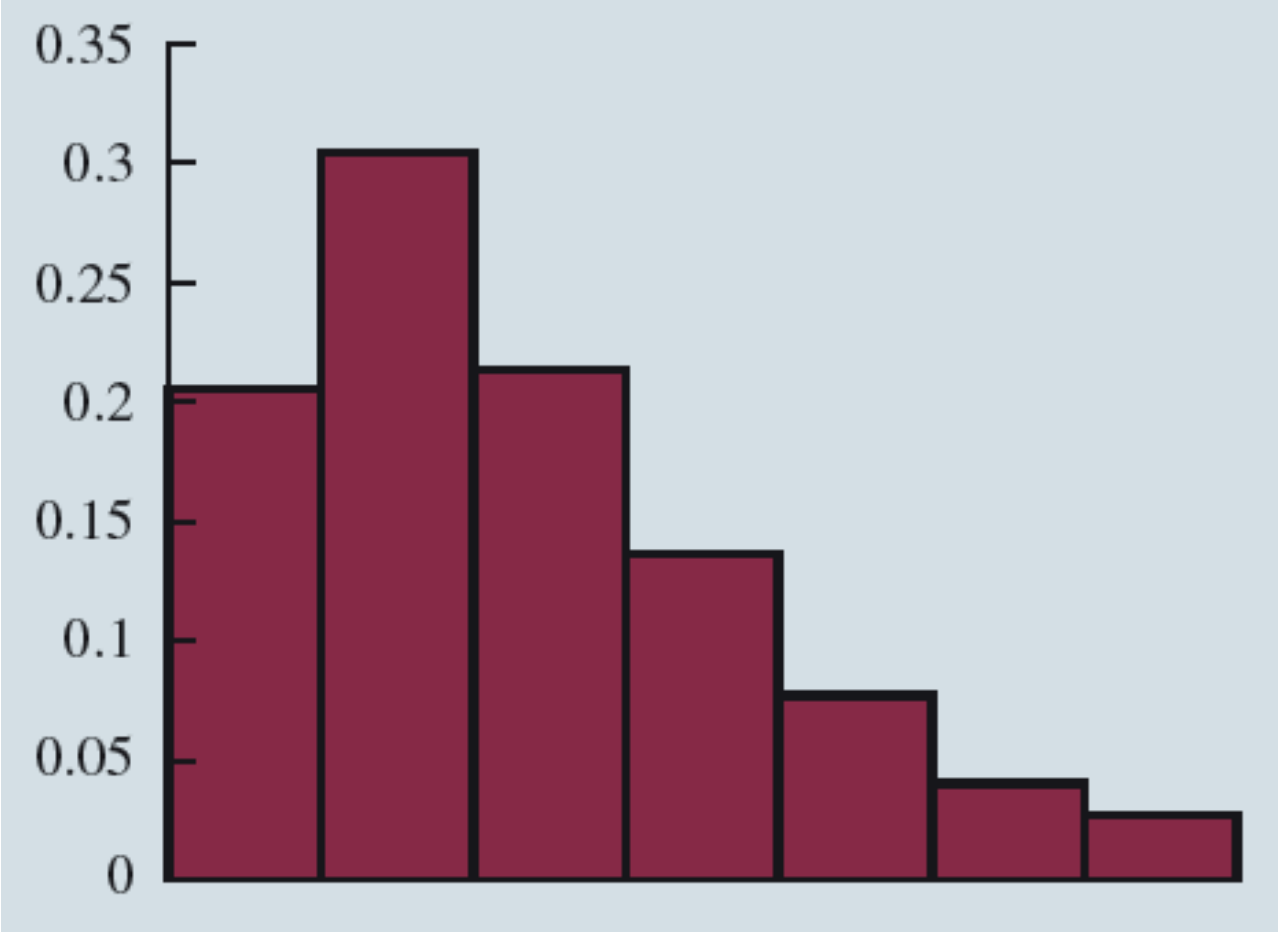
Histograms Showing Skewness (2 of 4)

- Moderately Skewed Left
  - A longer tail to the left
  - **Example:** Exam Scores



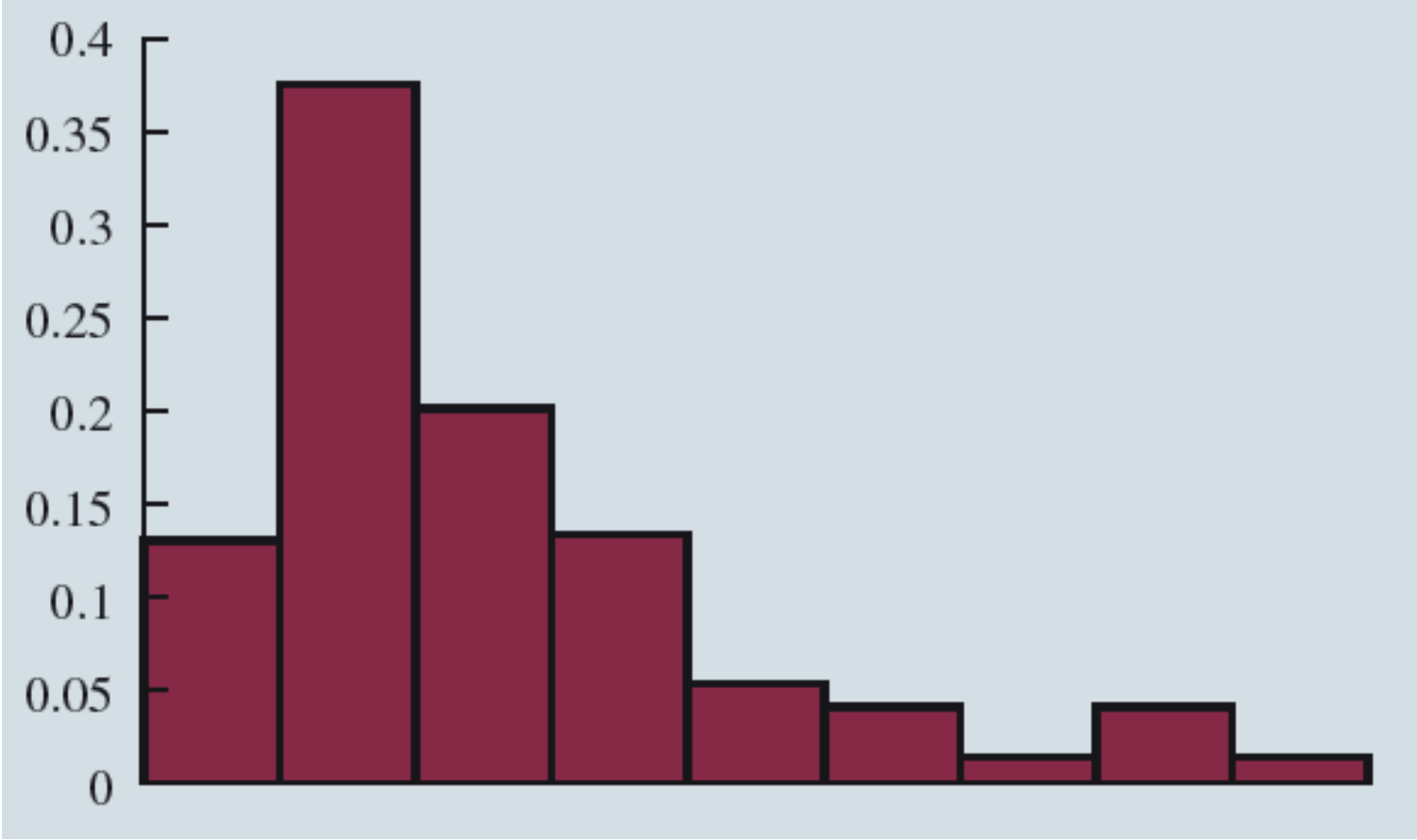
Histograms Showing Skewness (3 of 4)

- Moderately Right Skewed
  - A longer tail to the right
  - **Example:** Housing Values



Histograms Showing Skewness (4 of 4)

- Highly Skewed Right
  - A very long tail to the right
  - **Example:** Executive Salaries



Cumulative Distributions (1 of 3)

- Cumulative frequency distribution: Shows the *number* of items with values less than or equal to the upper limit of each class.
- Cumulative relative frequency distribution: Shows the *proportion* of items with values less than or equal to the upper limit of each class.
- Cumulative percent frequency distribution: Shows the *percentage* of items with values less than or equal to the upper limit of each class.

Cumulative Distributions (2 of 3)

- The last entry in a cumulative frequency distribution always equals the total number of observations.
- The last entry in a cumulative relative frequency distribution always equals 1.00.
- The last entry in a cumulative percent frequency distribution always equals 100.

Cumulative Distributions (3 of 3)

Example: Sanderson and Clifford

Audit time (Days)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percent Frequency
Less than or equal to 14	4	.20	20
Less than or equal to 19	12	.60	60
Less than or equal to 24	17	.85	85
Less than or equal to 29	19	.95	95
Less than or equal to 34	20	1.00	100

Stem-and-Leaf Display (1 of 7)

- A stem-and-leaf display shows both the rank order and shape of the distribution of the data.
- It is similar to a histogram on its side, but it has the advantage of showing the actual data values.
- The first digits of each data item are arranged to the left of a vertical line.
- To the right of the vertical line, we record the last digit for each item in rank order.
- Each line (row) in the display is referred to as a stem.
- Each digit on a stem is a leaf.

Stem-and-Leaf Display (2 of 7)

Example

The number of questions answered correctly on an aptitude test by 50 students is analyzed with the help of a stem-and-leaf display. The relevant data is given in the following table.

Stem-and-Leaf Display (3 of 7)

Number of questions answered correctly by 50 students

112	73	126	82	92	115	95	84	68	100
72	92	128	104	108	76	141	119	98	85
69	76	118	132	96	91	81	113	115	94
97	86	127	134	100	102	80	98	106	106
107	73	124	83	92	81	106	75	95	119

Stem-and-Leaf Display (4 of 7)

Stem	Leaf
6	9 8
7	2 3 6 3 6 5
8	6 2 3 1 1 0 4 5
9	7 2 2 6 2 1 5 8 8 5 4
10	7 4 8 0 2 6 6 0 6
11	2 8 5 9 3 5 9
12	6 8 7 4
13	2 4
14	1

Stretched Stem-and-Leaf Display (1 of 2)

If we believe the original stem-and-leaf display has condensed the data too much, we can stretch the display vertically by using two stems for each leading digit(s).

Whenever a stem value is stated twice, the first value corresponds to leaf values of 0 to 4, and the second value corresponds to leaf values of 5 to 9.

Stretched Stem-and-Leaf Display (2 of 2)

6	8	9				
7	2	3	3			
7	5	6	6			
8	0	1	1	2	3	4
8	5	6				
9	1	2	2	2	4	
9	5	5	6	7	8	8
10	0	0	2	4		
10	6	6	6	7	8	
11	2	3				
11	5	5	8	9	9	
12	4					
12	6	7	8			
13	2	4				
14	1					

Stem-and-Leaf Display (5 of 7)

Leaf Units

- A single digit is used to define each leaf.
- In the preceding example, the leaf unit was 1.
- Leaf units may be 100, 10, 1, 0.1, and so on.
- Where the leaf unit is not shown, it is assumed to equal 1.
- The leaf unit indicates how to multiply the stem-and-leaf numbers in order to approximate the original data.

Stem-and-Leaf Display (6 of 7)

Example: Leaf Unit = 0.1

If we have data with values such as

8.6 11.7 9.4 9.1 10.2 11.0 8.8

Leaf Unit = 0.1	
8	6 8
9	1 4
10	2
11	0 7

Stem-and-Leaf Display (7 of 7)

Example: Leaf Unit = 10

If we have data with values such as

1806 1717 1974 1791 1682 1910 1838

Leaf Unit = 10	
16	8
17	1 9
18	0 3
19	1 7

The 82 in 1682 is rounded down to 80 and is represented as an 8.