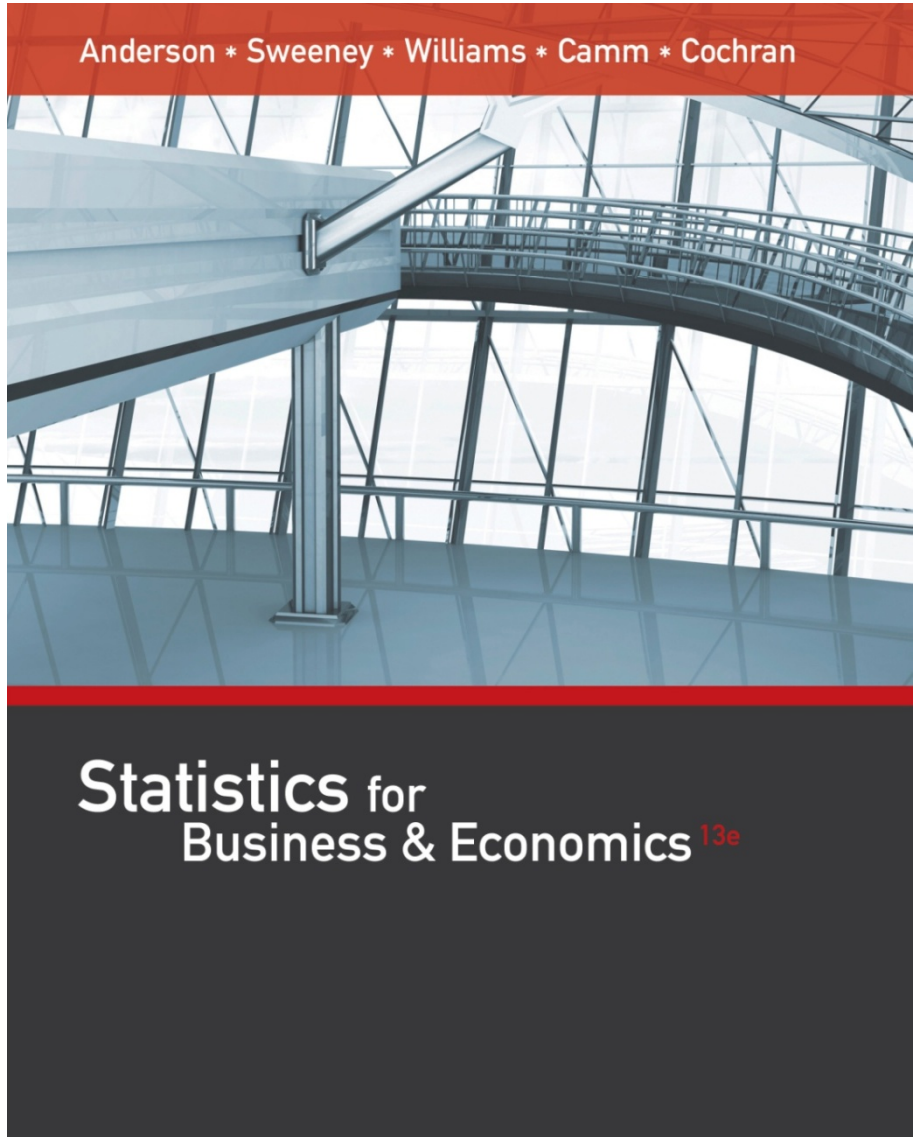
Statistics for Business and Economics (13e)

Anderson, Sweeney, Williams, Camm, Cochran

Slides by John Loucks

St. Edwards University



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 Categorical Data

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

Frequency Distribution

- 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

Example: Marada Inn

- Guests staying at Marada Inn were asked to rate the quality of their accommodations as being *excellent*, *above average*, *average*, *below average*, or *poor*.
- The ratings provided by a sample of 20 guests are:

Below Average	Average	Above Average
Above Average	Above Average	Above Average
Above Average	Below Average	Below Average
Average	Poor	Poor
Above Average	Excellent	Above Average
Average	Above Average	Average
Above Average	Average	

Frequency Distribution

- Example: Marada Inn

Rating	Frequency
Poor	2
Below Average	3
Average	5
Above Average	9
Excellent	1
Total	20

Relative Frequency Distribution

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

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

- A relative frequency distribution is a tabular summary 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

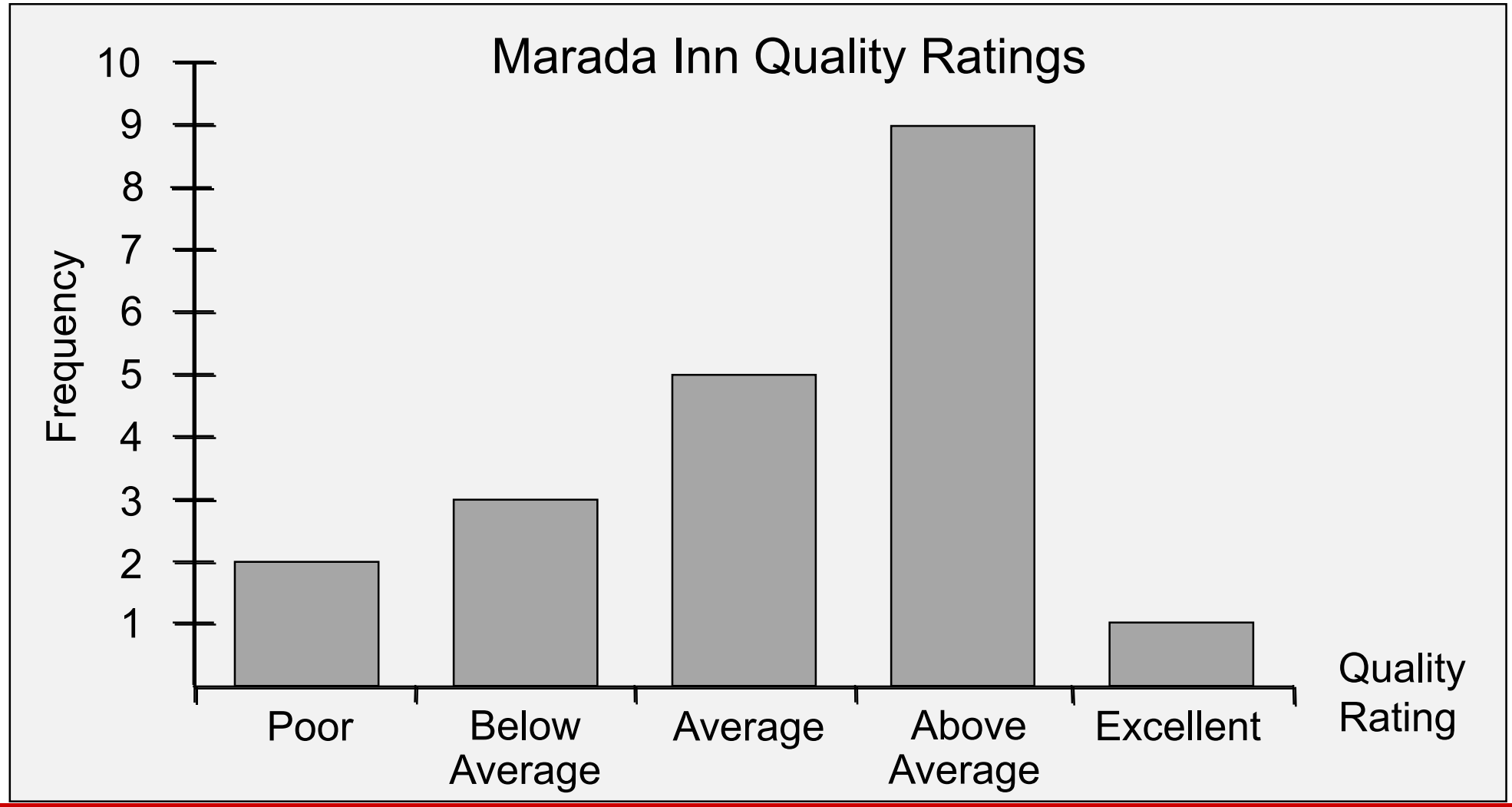
- Example: Marada Inn

Rating	Relative Frequency	Percent Frequency
Poor	.10	10
Below Average	.15	15
Average	.25	25
Above Average	.45	45
Excellent	.05	5
Total	1.00	100

Bar Chart

- A bar chart is a graphical display for depicting qualitative 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 a separate category.

Bar Chart



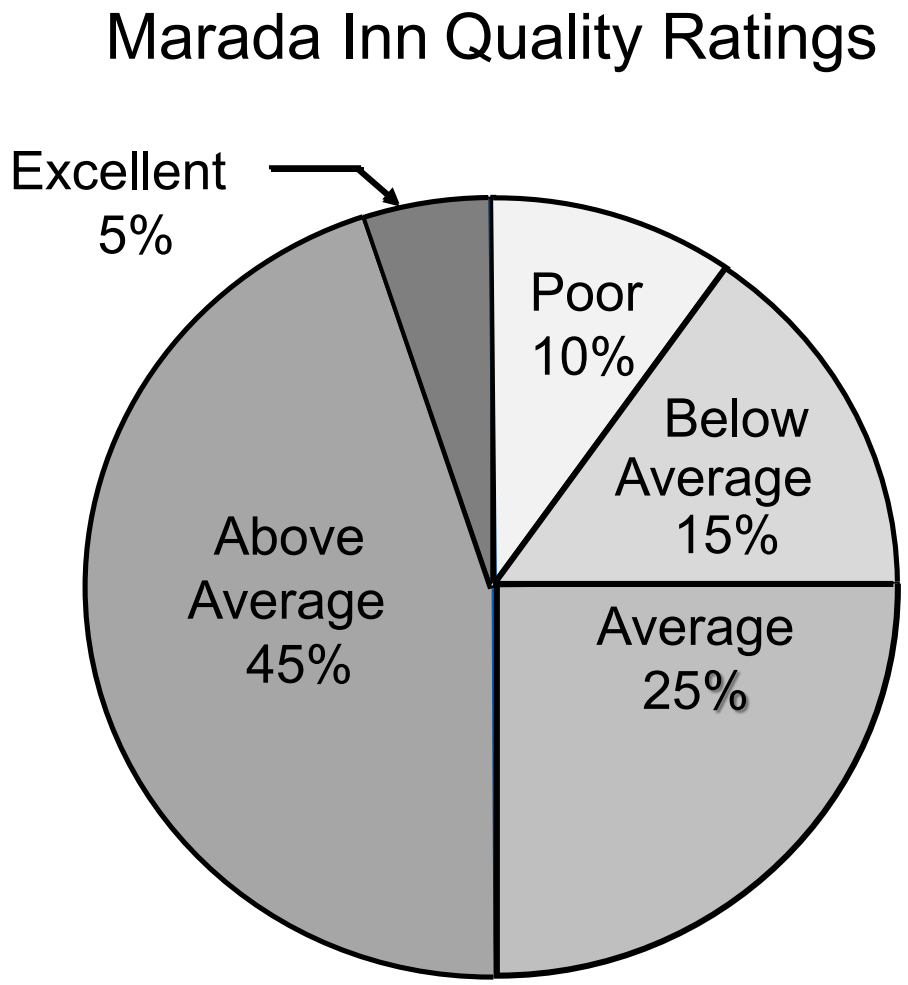
Pareto Diagram

- In quality control, bar charts are used to identify the most important causes of problems.
- When the bars are arranged in descending order of height from left to right (with the most frequently occurring cause appearing first) the bar chart is called a Pareto diagram.
- This diagram is named for its founder, Vilfredo Pareto, an Italian economist.

Pie Chart

- 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



Example: Marada Inn

- One-half of the customers surveyed gave Marada a quality rating of “above average” or “excellent” (looking at the left side of the pie). This might please the manager.
- For each customer who gave an “excellent” rating, there were two customers who gave a “poor” rating (looking at the top of the pie). This should displease the manager.

Summarizing Quantitative Data

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

Frequency Distribution

- Example: Hudson Auto Repair

  The manager of Hudson Auto would like to gain a better understanding of the cost of parts used in the engine tune-ups performed in the shop. She examines 50 customer invoices for tune-ups. The costs of parts, rounded to the nearest dollar, are listed on the next slide.

Frequency Distribution

- Example: Hudson Auto Repair

Sample of Parts Cost($) for 50 Tune-ups

91	78	93	57	75	52	99	80	97	62
71	69	72	89	66	75	79	75	72	76
104	74	62	68	97	105	77	65	80	109
85	97	88	68	83	68	71	69	67	74
62	82	98	101	79	105	79	69	62	73

Frequency Distribution

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

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

Frequency Distribution

- 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

- Guidelines for Determining the Width of Each Class
  - Use classes of equal width.
  - 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

- 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

- 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

- 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

- Example: Hudson Auto Repair

  If we choose six classes:

  Approximate Class Width = (109 - 50)/6 = 9.83≅ 10

Parts Cost ($)	Frequency
50-59	2
60-69	13
70-79	16
80-89	7
90-99	7
100-109	5
Total	50

Relative Frequency and Percent Frequency Distributions

- Example: Hudson Auto Repair

Parts Cost ($)	Relative Frequency	Percent Frequency
50-59	.04 = 2/50	4 = .04(100)
60-69	.26	26
70-79	.32	32
80-89	.14	14
90-99	.14	14
100-109	.10	10
Total	1.00	100

Relative Frequency and Percent Frequency Distributions

- Example: Hudson Auto Repair

  Insights Gained from the Percent Frequency Distribution:

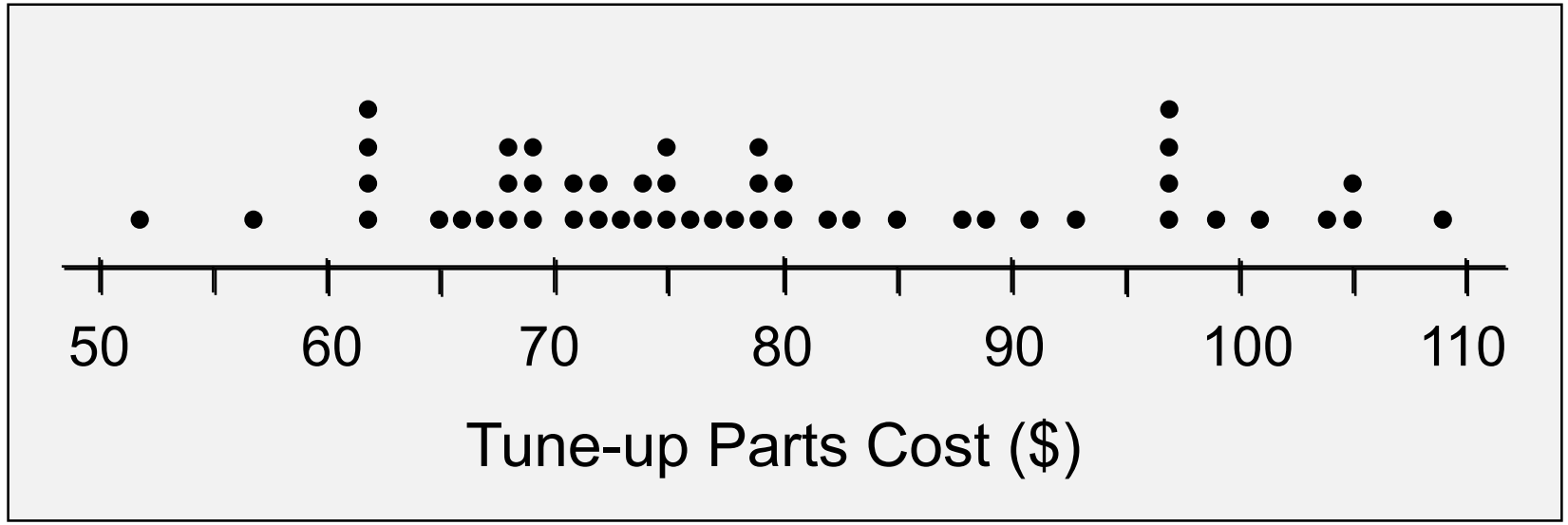
  - Only 4% of the parts costs are in the $50-59 class.
  - 30% of the parts costs are under $70.
  - The greatest percentage (32% or almost one-third) of the parts costs are in the $70-79 class.
  - 10% of the parts costs are $100 or more.

Dot Plot

- 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

- Example: Hudson Auto Repair

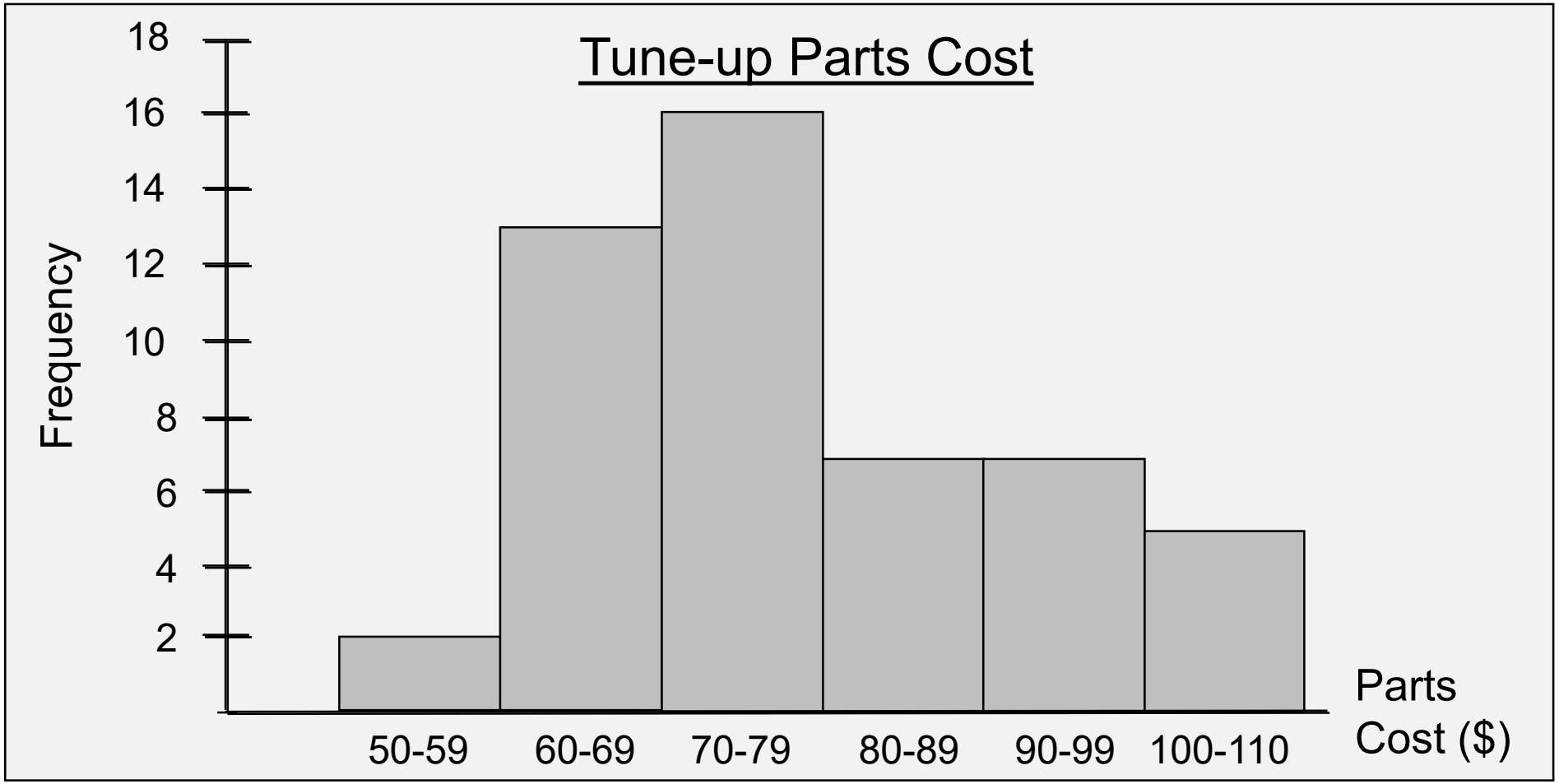


Histogram

- 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.

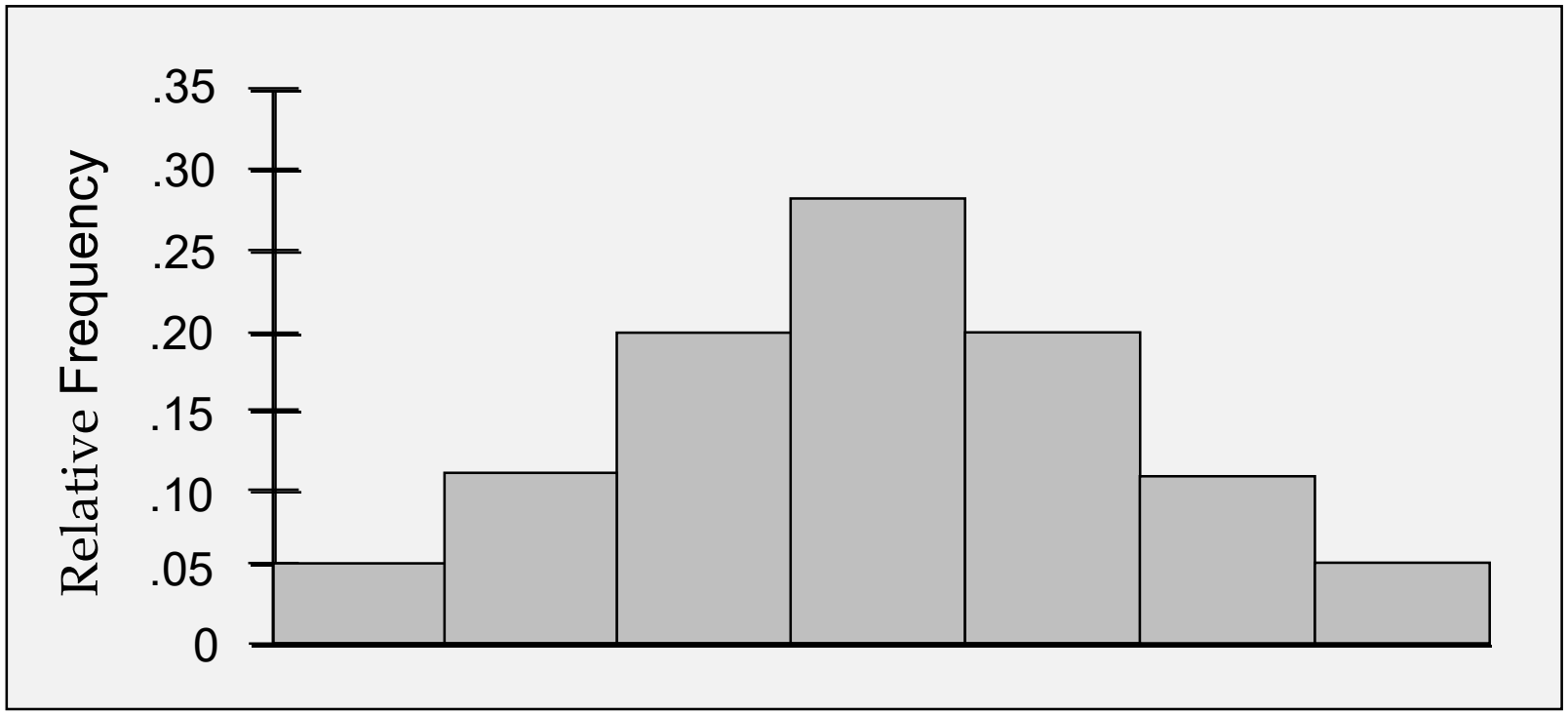
Histogram

- Example: Hudson Auto Repair



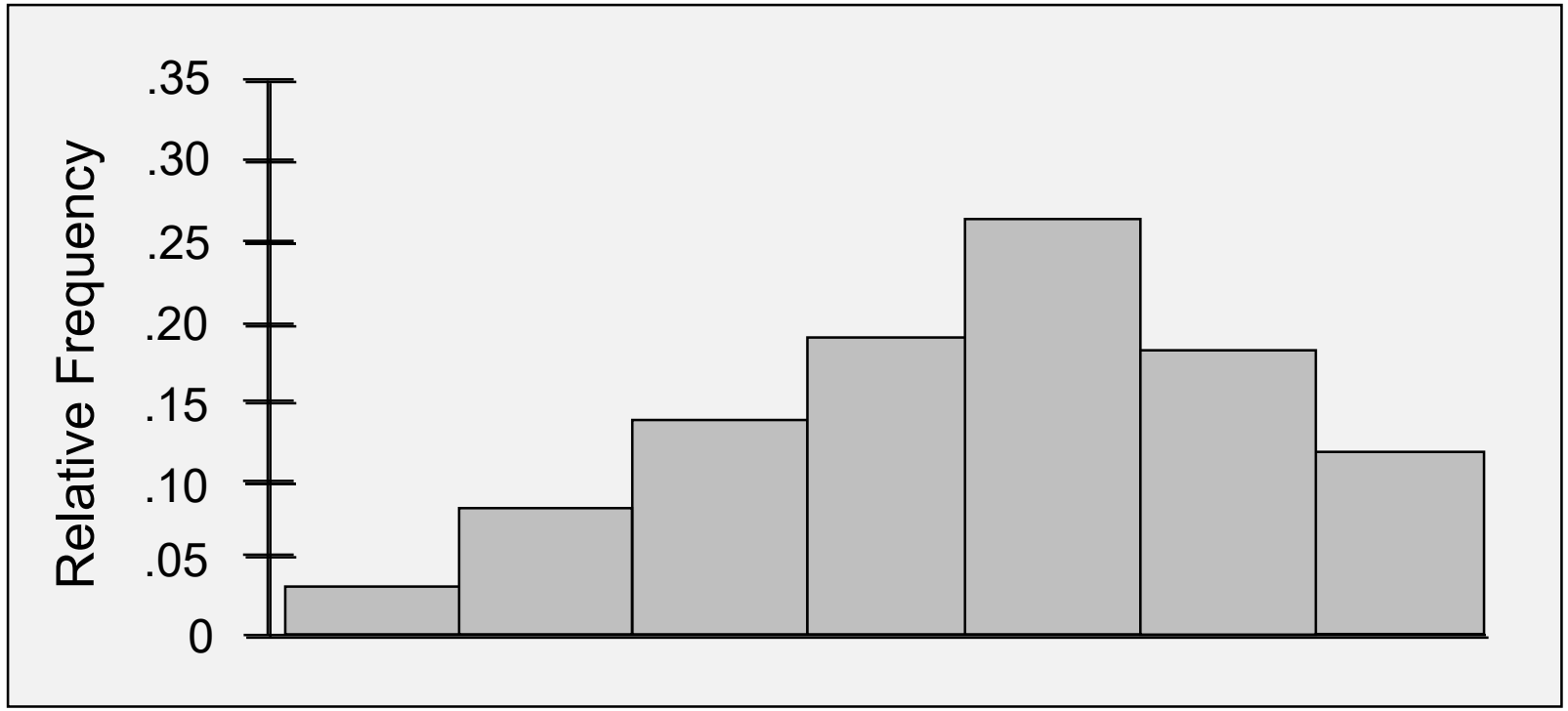
Histograms Showing Skewness

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



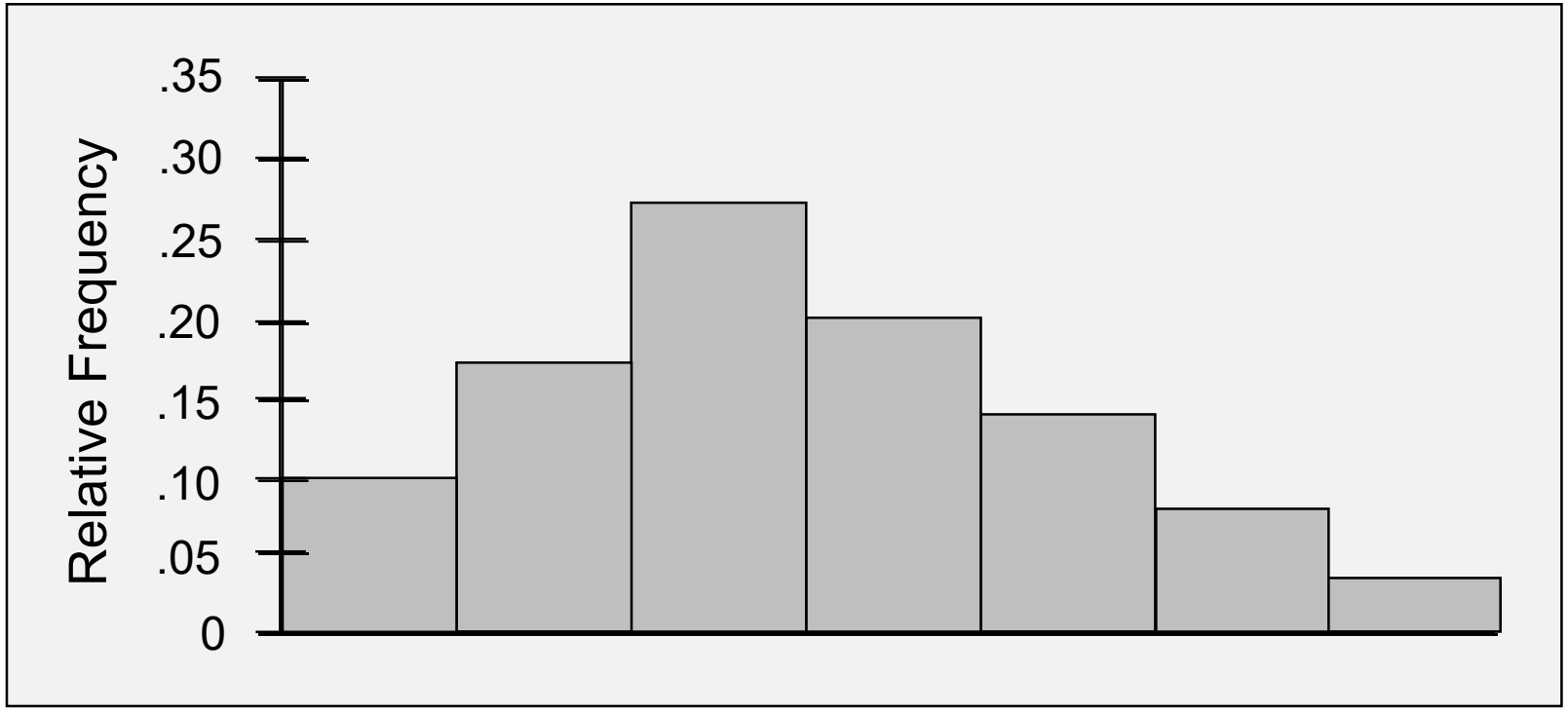
Histograms Showing Skewness

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



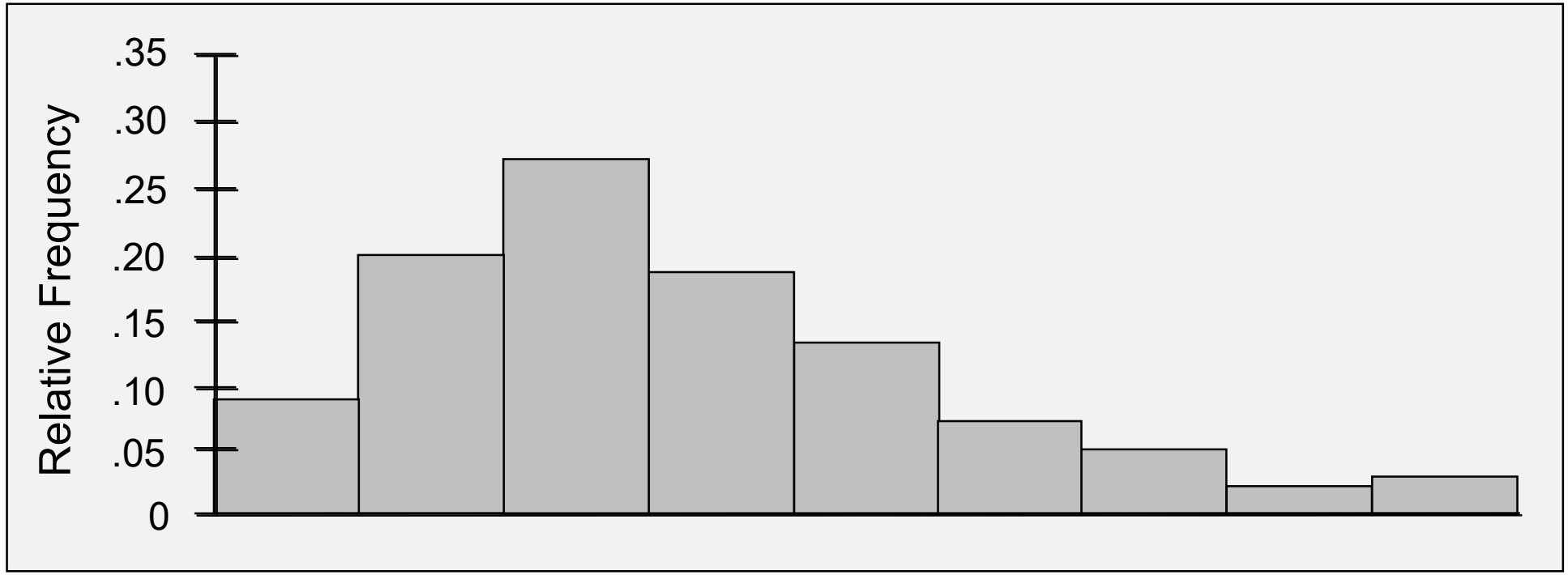
Histograms Showing Skewness

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



Histograms Showing Skewness

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



Cumulative Distributions

- 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

- 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

- Hudson Auto Repair

Cost ($)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percent Frequency
≤ 59	2	.04	4
≤ 69	15 = 2+13	.30 = 15/50	30 = .30(100)
≤ 79	31	.62	62
≤ 89	38	.76	76
≤ 99	45	.90	90
≤ 109	50	1.00	100

Stem-and-Leaf Display

- A stem-and-leaf display shows both the rank order and shape of a distribution of data.
- It is similar to a histogram on its side, but it has the advantage of showing the actual data values.
- The leading 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.

Example: Hudson Auto Repair

The manager of Hudson Auto would like to gain a better understanding of the cost of parts used in the engine tune-ups performed in the shop. She examines 50 customer invoices for tune-ups. The costs of parts, rounded to the nearest dollar, are listed on the next slide.

Stem-and-Leaf Display

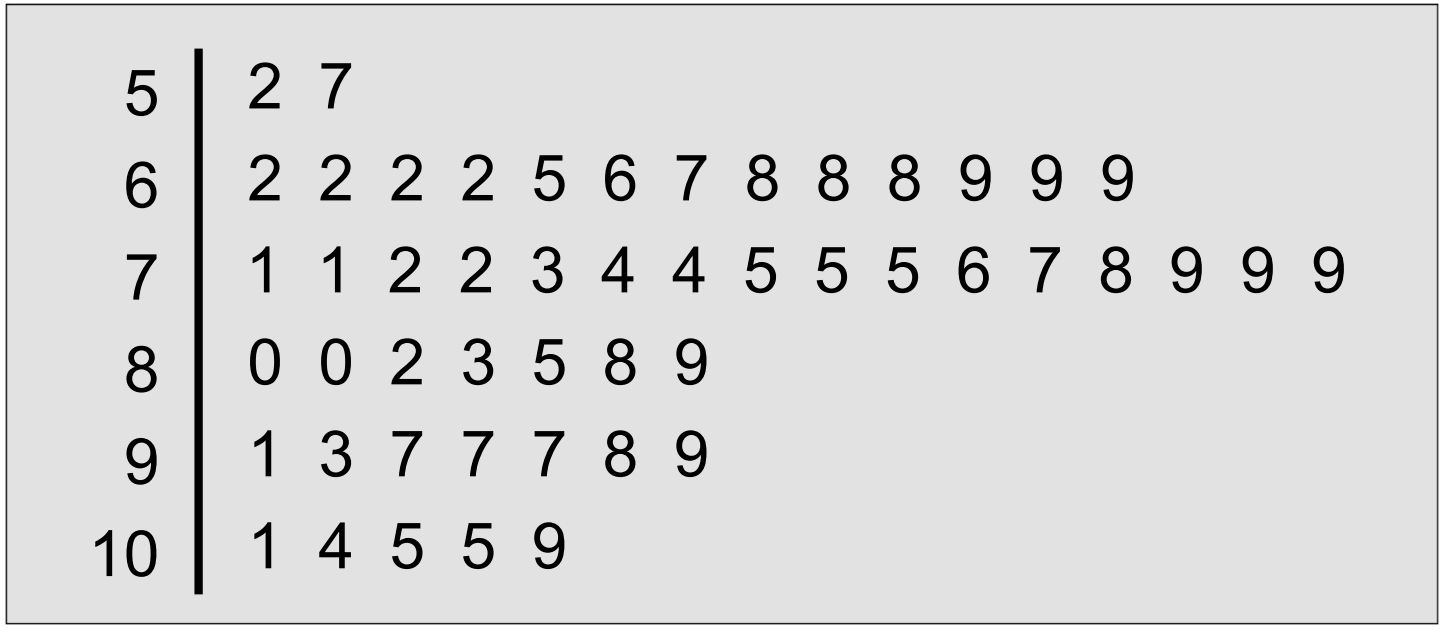
- Example: Hudson Auto Repair

Sample of Parts Cost ($) for 50 Tune-ups

91	78	93	57	75	52	99	80	97	62
71	69	72	89	66	75	79	75	72	76
104	74	62	68	97	105	77	65	80	109
85	97	88	68	83	68	71	69	67	74
62	82	98	101	79	105	79	69	62	73

Stem-and-Leaf Display

- Example: Hudson Auto Repair



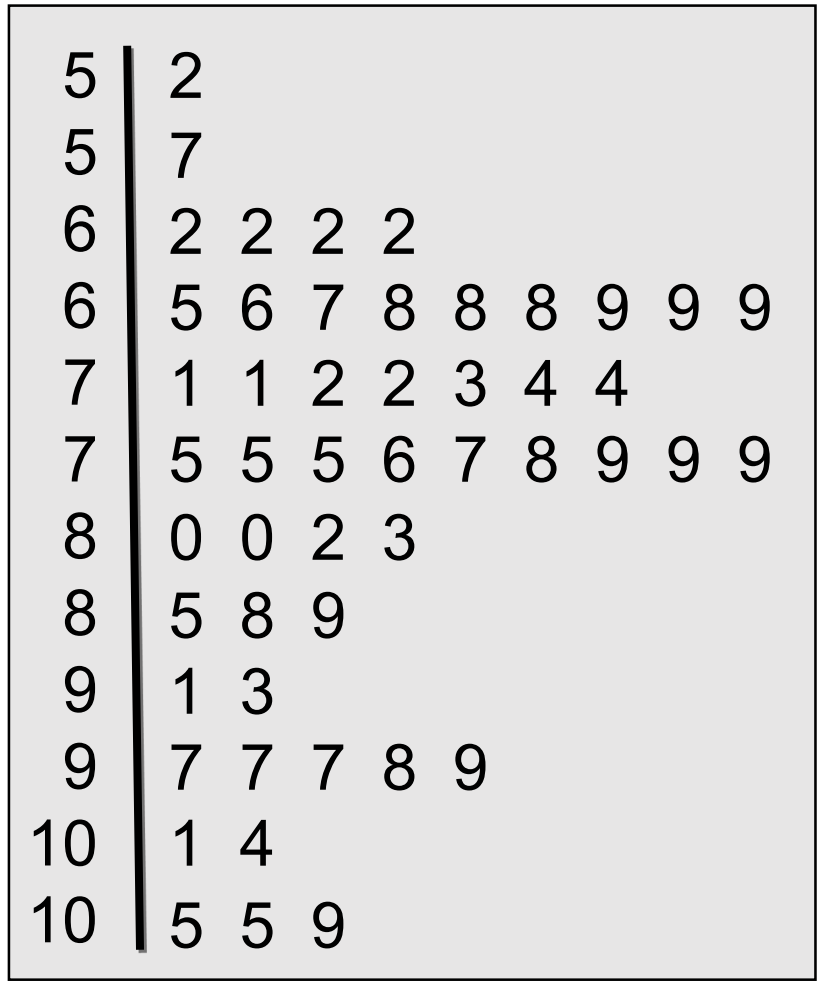
Stems Leaves

Stretched Stem-and-Leaf Display

- 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 - 4, and the second value corresponds to leaf values of 5 - 9.

Stretched Stem-and-Leaf Display

- Example: Hudson Auto Repair



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

Stem-and-Leaf Display

- 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

- 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

- 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.

End of Chapter 2, Part A