

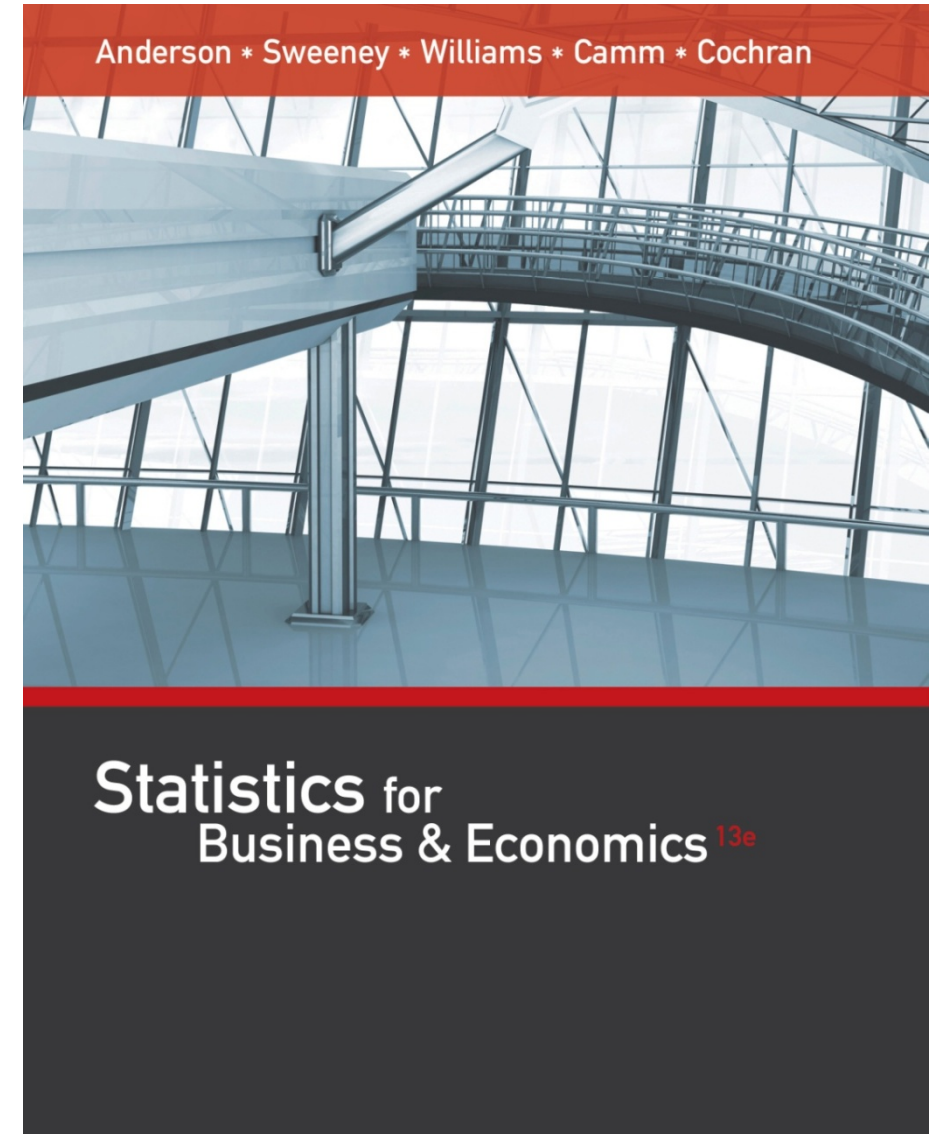
Essentials of Modern Business Statistics (6e)

Anderson, Sweeney, Williams, Camm, Cochran

© 2017 Cengage Learning

Slides by John Loucks

St. Edwards University



Chapter 2, Part B

Descriptive Statistics: Tabular and Graphical Displays

- Summarizing Data for Two Variables Using Tables
- Summarizing Data for Two Variables Using Graphical Displays
- Data Visualization: Best Practices in Creating Effective Graphical Displays

Summarizing Data for Two Variables using Tables

- Thus far we have focused on methods that are used to summarize the data for one variable at a time.
- Often a manager is interested in tabular and graphical methods that will help understand the relationship between two variables.
- Crosstabulation is a method for summarizing the data for two variables.

Crosstabulation

- A crosstabulation is a tabular summary of data for two variables.
- Crosstabulation can be used when:
 - one variable is categorical and the other is quantitative,
 - both variables are categorical, or
 - both variables are quantitative.
- The left and top margin labels define the classes for the two variables.

Crosstabulation

- Example: Finger Lakes Homes

The number of Finger Lakes homes sold for each style and price for the past two years is shown below.

Price Range	<u>Home Style</u>				Total
	Colonial	Log	Split	A-Frame	
< \$250,000	18	6	19	12	55
≥ \$250,000	12	14	16	3	45
Total	30	20	35	15	100

Crosstabulation

- Example: Finger Lakes Homes

Insights Gained from Preceding Crosstabulation

- The greatest number of homes (19) in the sample are a split-level style and priced at less than \$250,000.
- Only three homes in the sample are an A-Frame style and priced at \$250,000 or more.

Crosstabulation: Row or Column Percentages

- Converting the entries in the table into row percentages or column percentages can provide additional insight about the relationship between the two variables.

Crosstabulation: Row Percentages

- Example: Finger Lakes Homes

Price Range	<u>Home Style</u>				Total
	Colonial	Log	Split	A-Frame	
< \$250,000	32.73	10.91	34.55	21.82	100
≥ \$250,000	26.67	31.11	35.56	6.67	100

Note: row totals are actually 100.01 due to rounding.



$$(\text{Colonial and } \geq \$250\text{K}) / (\text{All } \geq \$250\text{K}) \times 100 = (12/45) \times 100$$

Crosstabulation: Column Percentages

- Example: Finger Lakes Homes

Price Range	<u>Home Style</u>			
	Colonial	Log	Split	A-Frame
< \$250,000	60.00	30.00	54.29	80.00
≥ \$250,000	40.00	70.00	45.71	20.00
Total	100	100	100	100



$$(\text{Colonial and } \geq \$250\text{K}) / (\text{All Colonial}) \times 100 = (12/30) \times 100$$

Crosstabulation: Simpson's Paradox

- Data in two or more crosstabulations are often aggregated to produce a summary crosstabulation.
- We must be careful in drawing conclusions about the relationship between the two variables in the aggregated crosstabulation.
- In some cases the conclusions based upon an aggregated crosstabulation can be completely reversed if we look at the unaggregated data. The reversal of conclusions based on aggregate and unaggregated data is called Simpson's paradox.

Summarizing Data for Two Variables

Using Graphical Displays

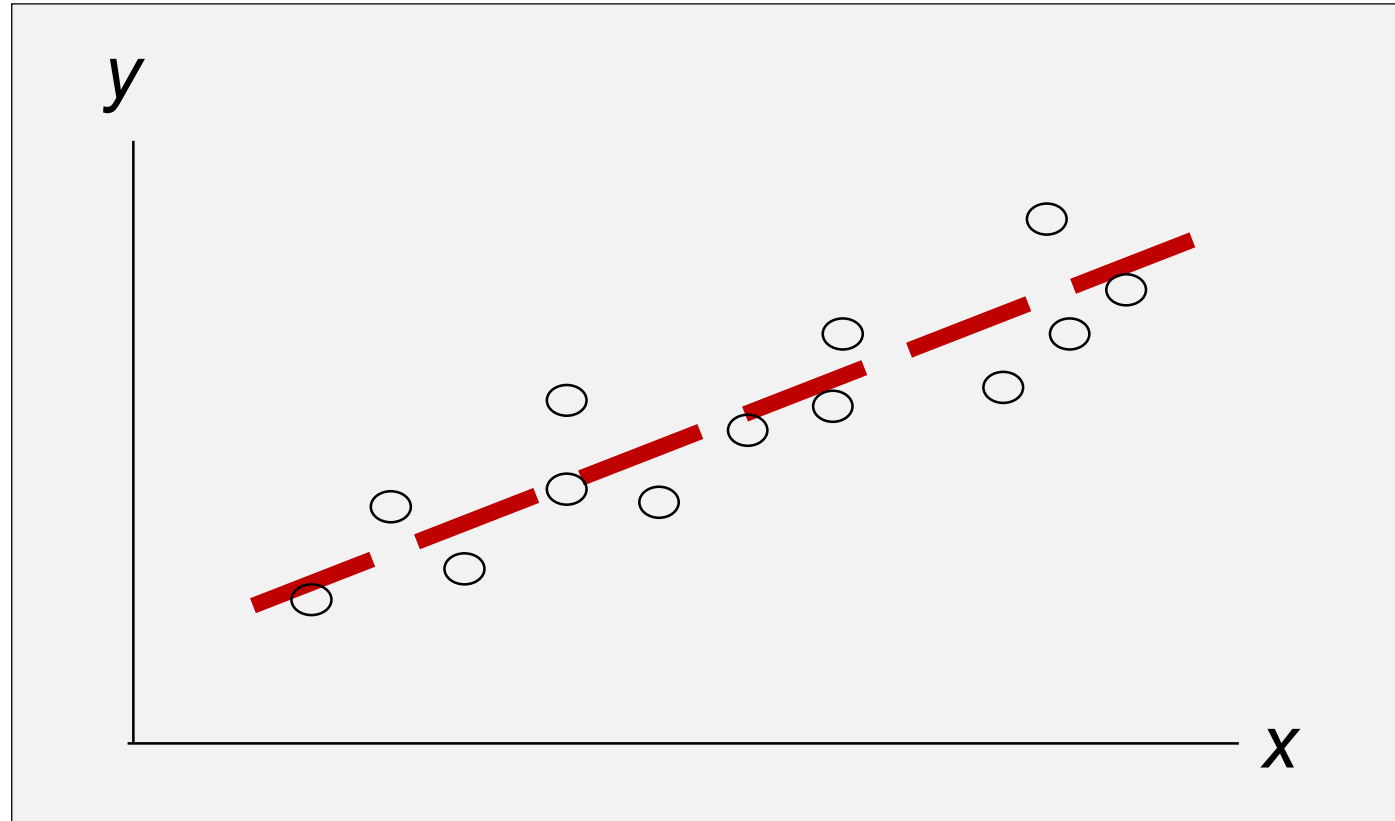
- In most cases, a graphical display is more useful than a table for recognizing patterns and trends.
- Displaying data in creative ways can lead to powerful insights.
- Scatter diagrams and trendlines are useful in exploring the relationship between two variables.

Scatter Diagram and Trendline

- A scatter diagram is a graphical presentation of the relationship between two quantitative variables.
- One variable is shown on the horizontal axis and the other variable is shown on the vertical axis.
- The general pattern of the plotted points suggests the overall relationship between the variables.
- A trendline provides an approximation of the relationship.

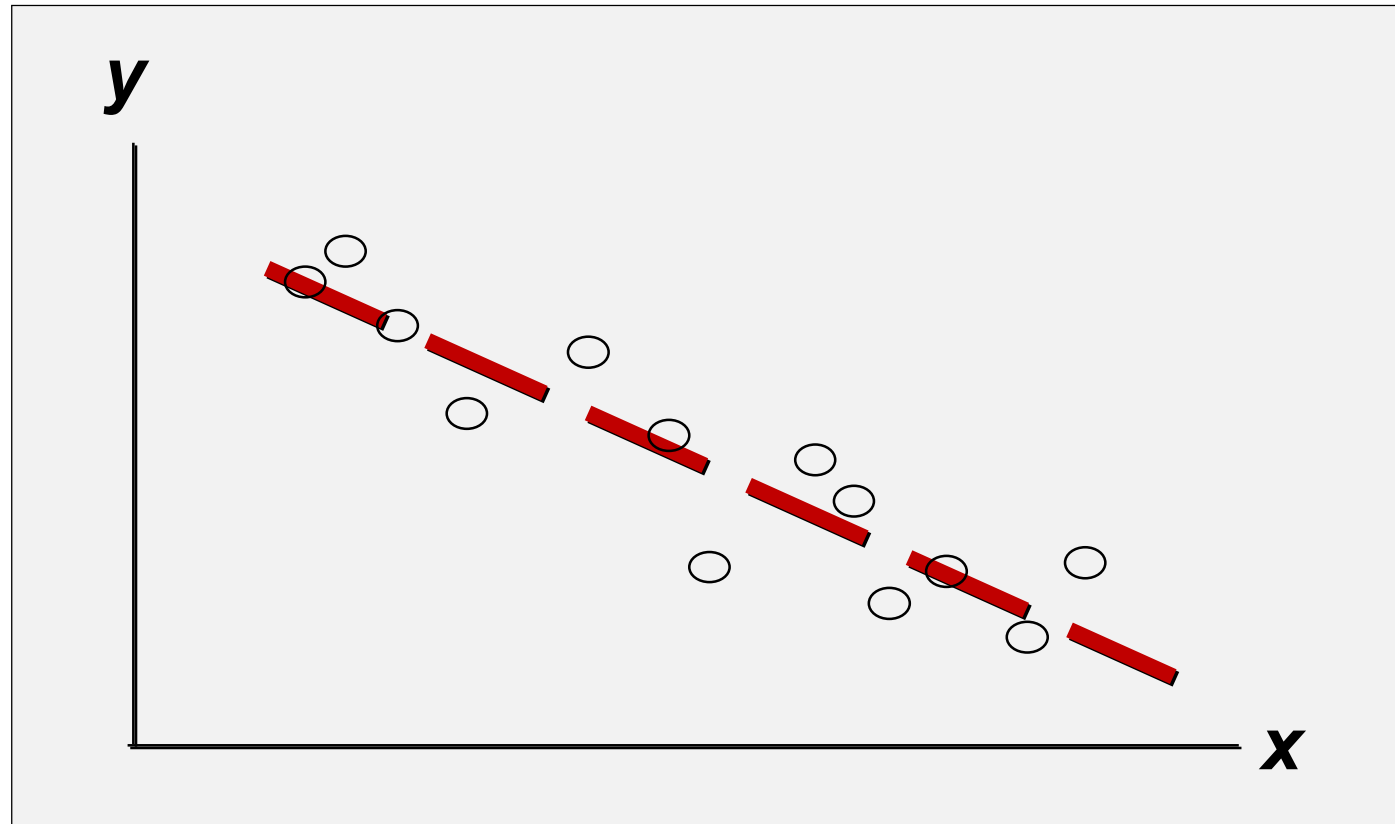
Scatter Diagram

- A Positive Relationship



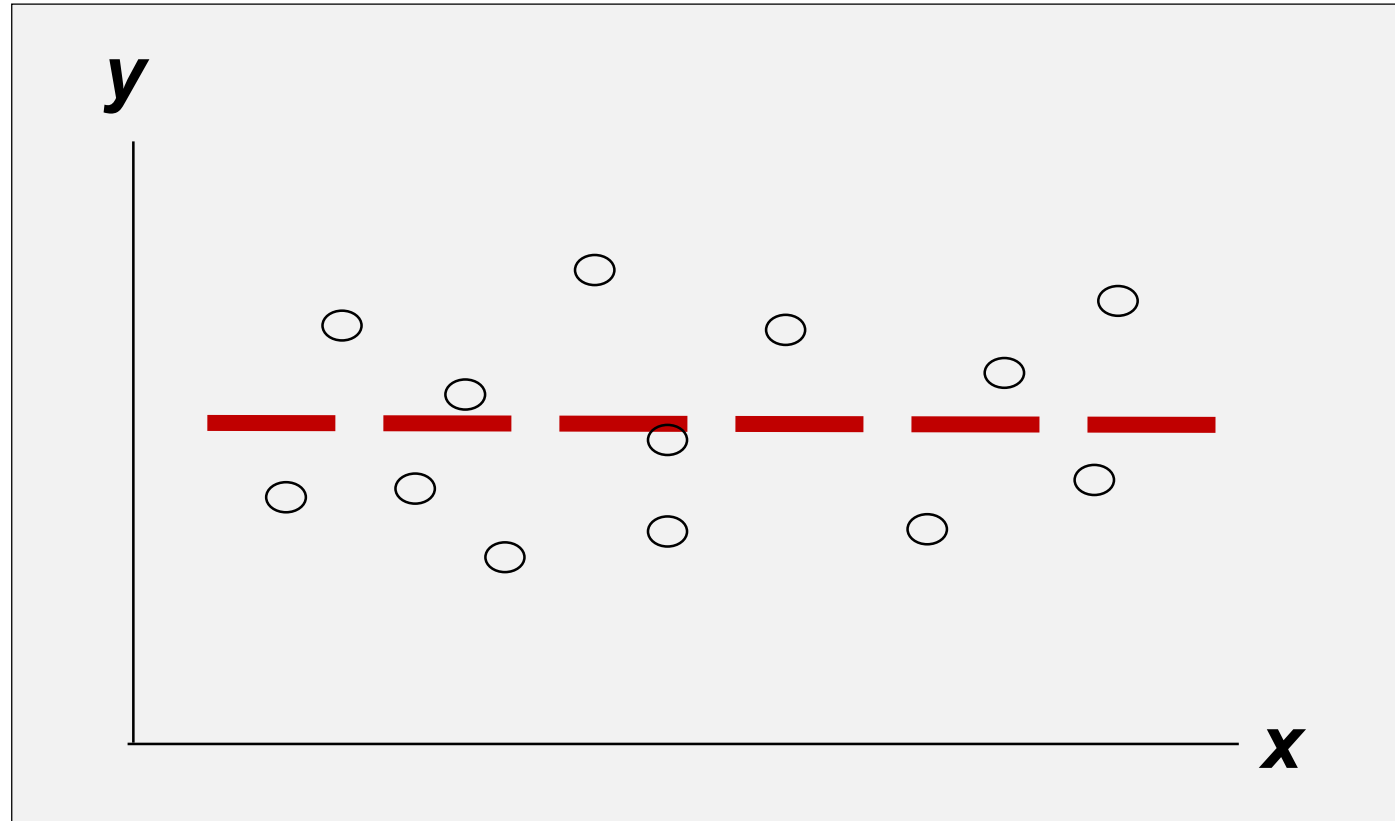
Scatter Diagram

- A Negative Relationship



Scatter Diagram

- No Apparent Relationship



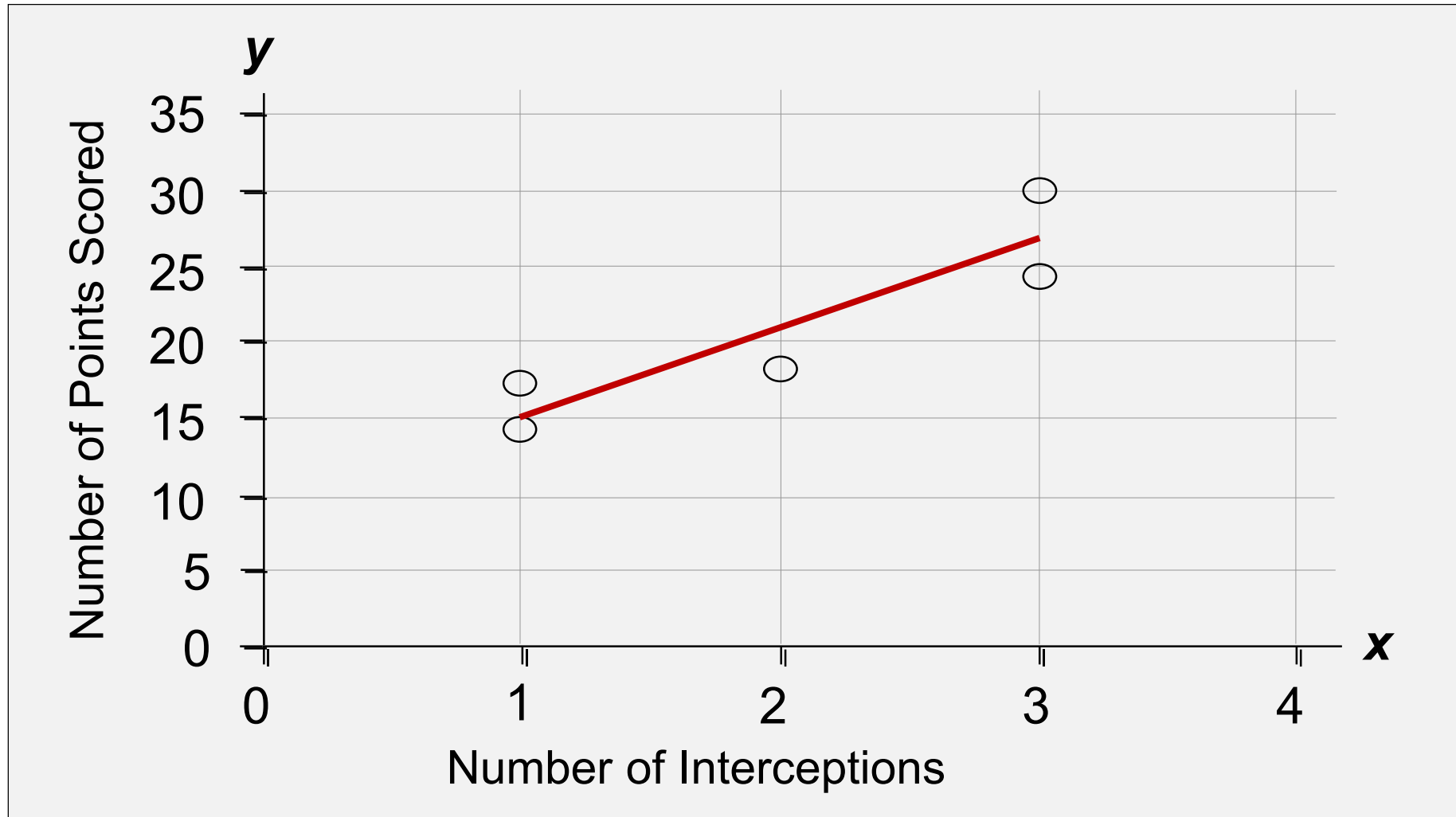
Scatter Diagram

- Example: Panthers Football Team

The Panthers football team is interested in investigating the relationship, if any, between interceptions made and points scored.

x = Number of <u>Interceptions</u>	y = Number of <u>Points Scored</u>
1	14
3	24
2	18
1	17
3	30

Scatter Diagram and Trendline



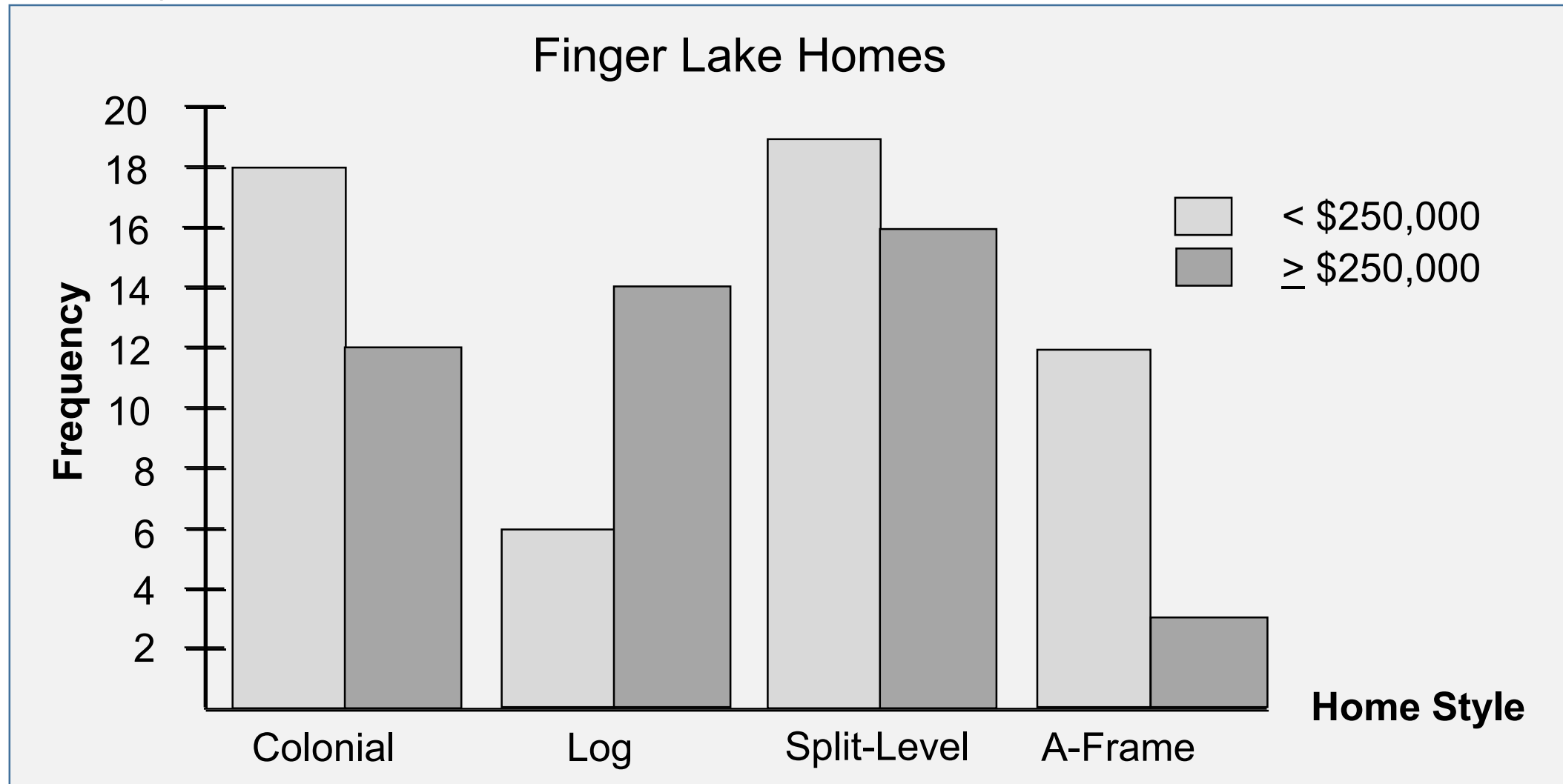
Example: Panthers Football Team

- Insights Gained from the preceding Scatter Diagram
 - The scatter diagram indicates a positive relationship between the number of interceptions and the number of points scored.
 - Higher points scored are associated with a higher number of interceptions.
 - The relationship is not perfect; all plotted points in the scatter diagram are not on a straight line.

Side-by-Side Bar Chart

- A side-by-side bar chart is a graphical display for depicting multiple bar charts on the same display.
- Each cluster of bars represents one value of the first variable.
- Each bar within a cluster represents one value of the second variable.

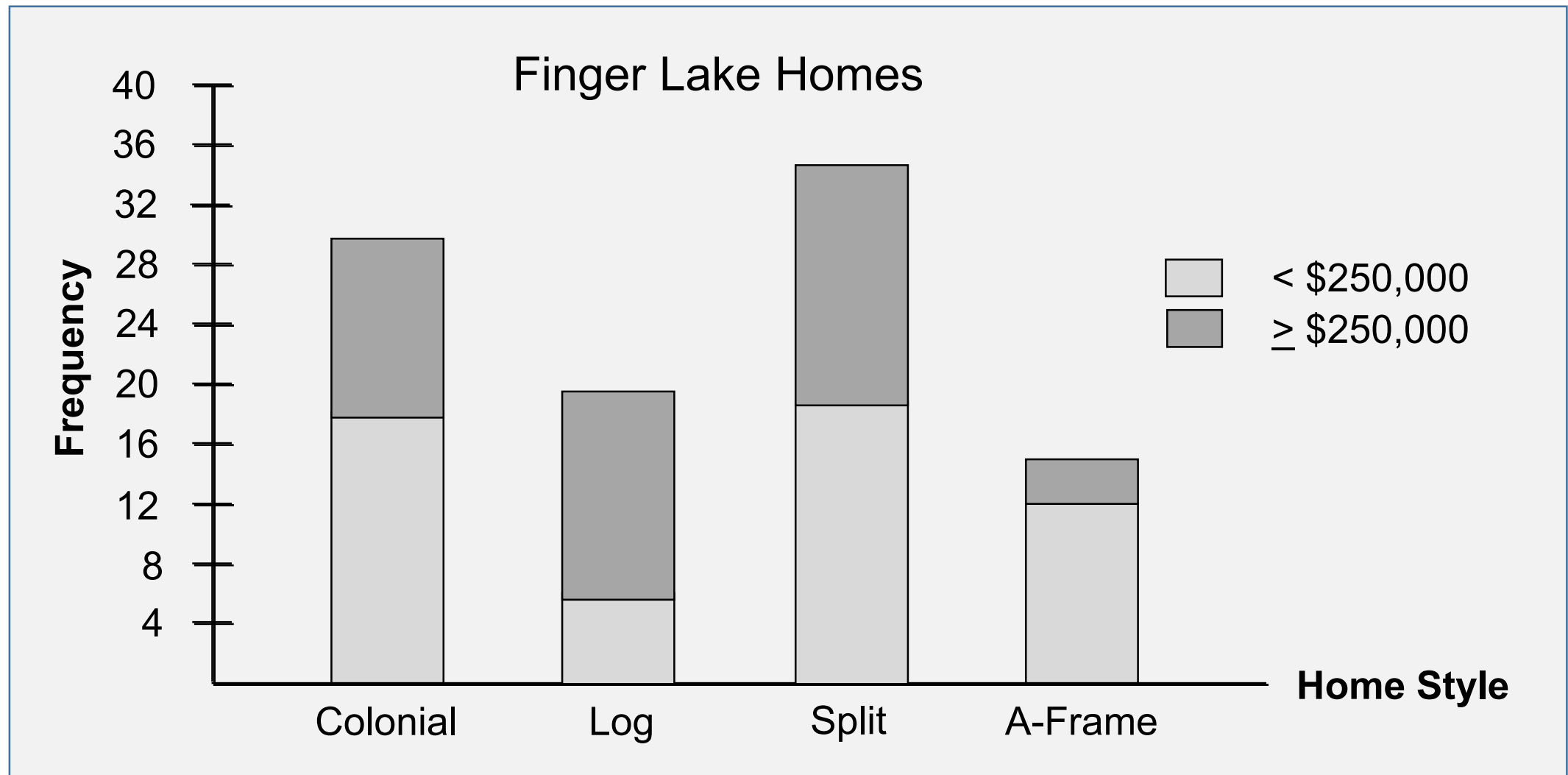
Side-by-Side Bar Chart



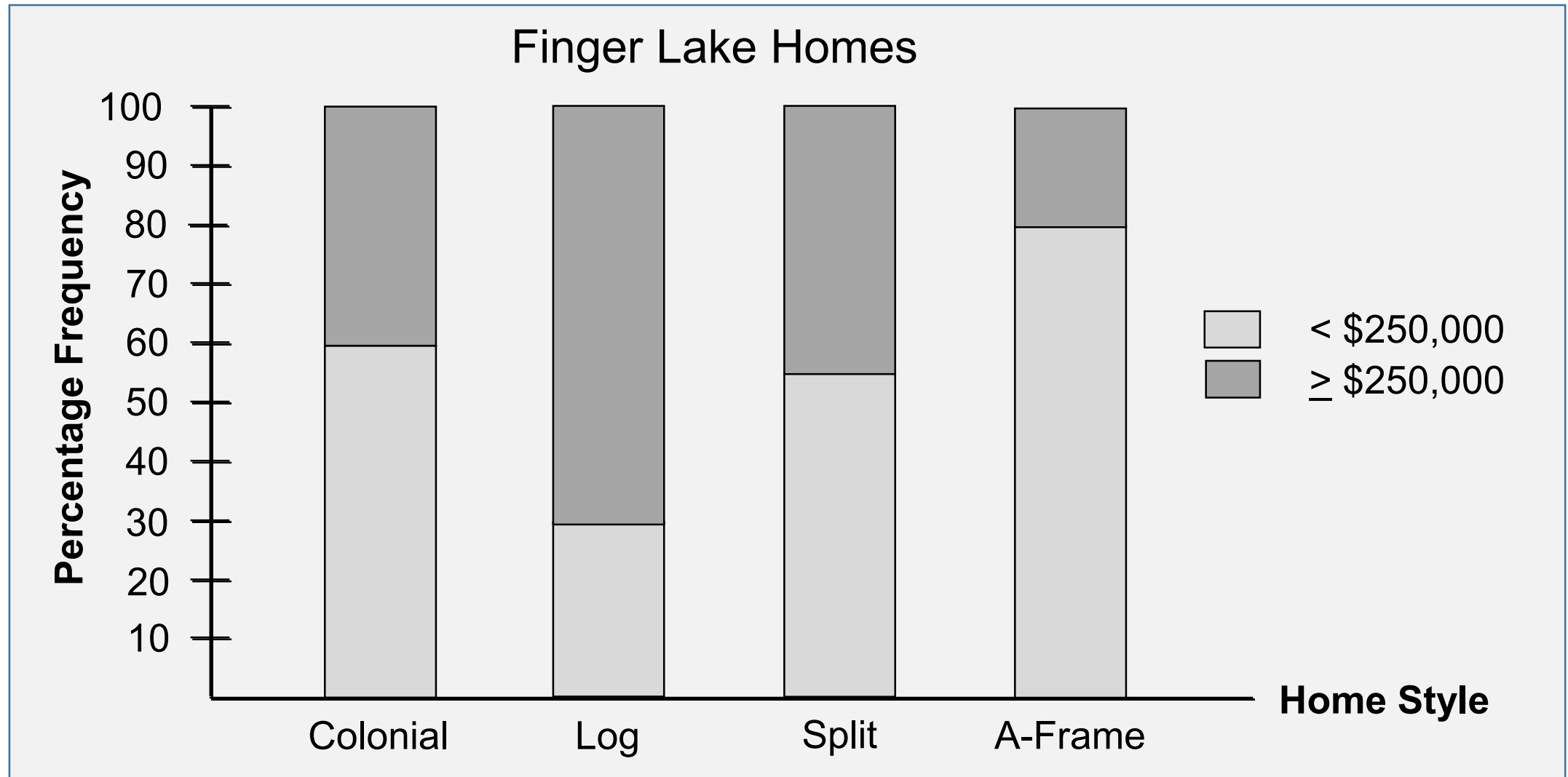
Stacked Bar Chart

- A stacked bar chart is another way to display and compare two variables on the same display.
- It is a bar chart in which each bar is broken into rectangular segments of a different color.
- If percentage frequencies are displayed, all bars will be of the same height (or length), extending to the 100% mark.

Stacked Bar Chart



Stacked Bar Chart



Data Visualization: Best Practices in Creating Effective Graphical Displays

- Data visualization is the use of graphical displays to summarize and present information about a data set.
- The goal is to communicate as effectively and clearly as possible, the key information about the data.

Creating Effective Graphical Displays

- Creating effective graphical displays is as much art as it is science.
- Here are some guidelines . . .
 - Give the display a clear and concise title.
 - Keep the display simple.
 - Clearly label each axis and provide the units of measure.
 - If colors are used, make sure they are distinct.
 - If multiple colors or line types are used, provide a legend.

Choosing the Type of Graphical Display

- Displays used to show the distribution of data:

Bar Chart to show the frequency distribution and relative frequency distribution for categorical data

Pie Chart to show the relative frequency and percent frequency for categorical data

Dot Plot to show the distribution for quantitative data over the entire range of the data

Histogram to show the frequency distribution for quantitative data over a set of class intervals

Stem-and-Leaf Display to show both the rank order and shape of the distribution for quantitative data

Choosing the Type of Graphical Display

- Displays used to make comparisons:

Side-by-Side Bar Chart to compare two variables

Stacked Bar Chart to compare the relative frequency or percent frequency of two categorical variables

Choosing the Type of Graphical Display

- Displays used to show relationships:

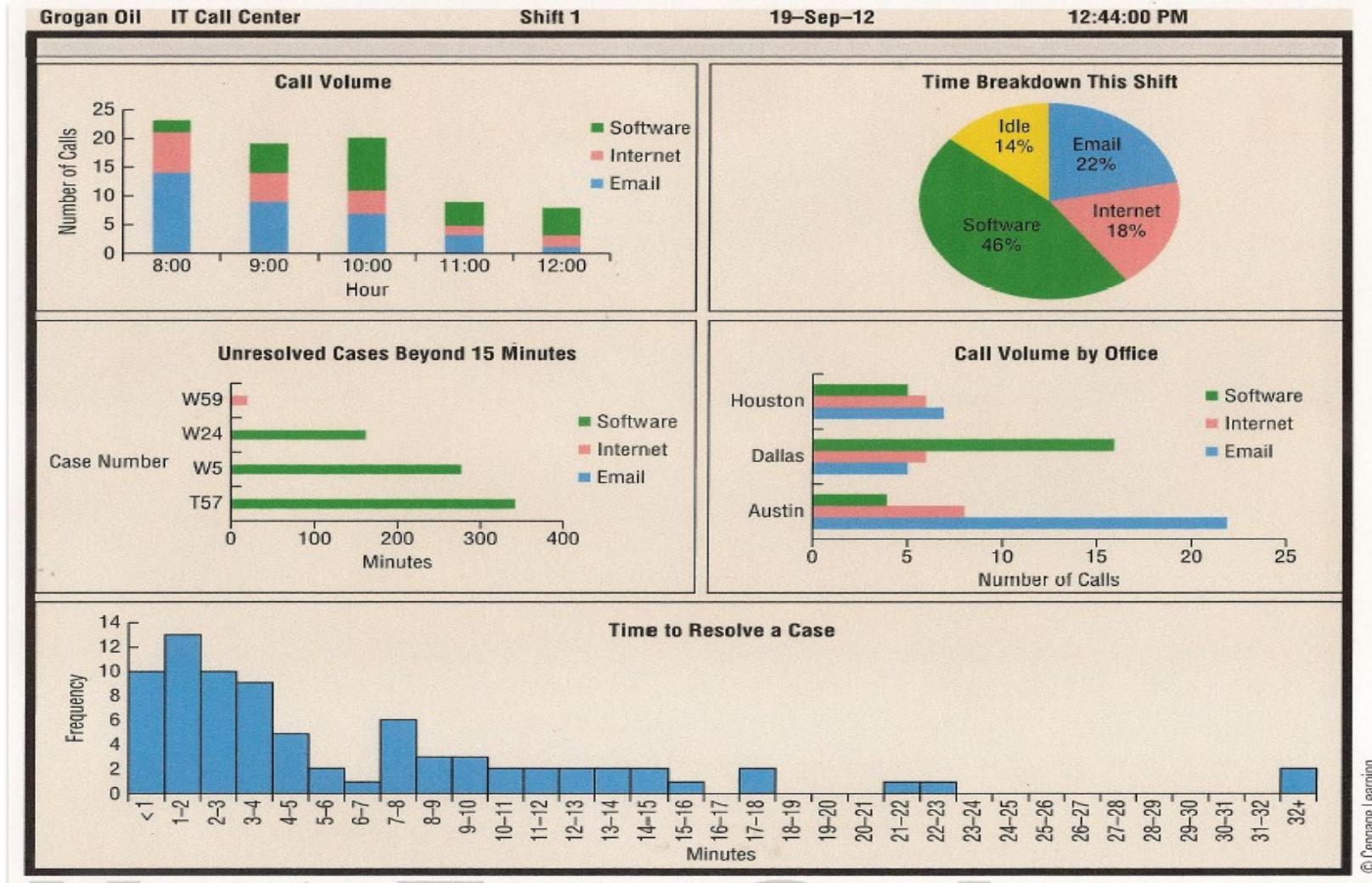
Scatter Diagram to show the relationship between two quantitative variables

Trendline to approximate the relationship of data in a scatter diagram

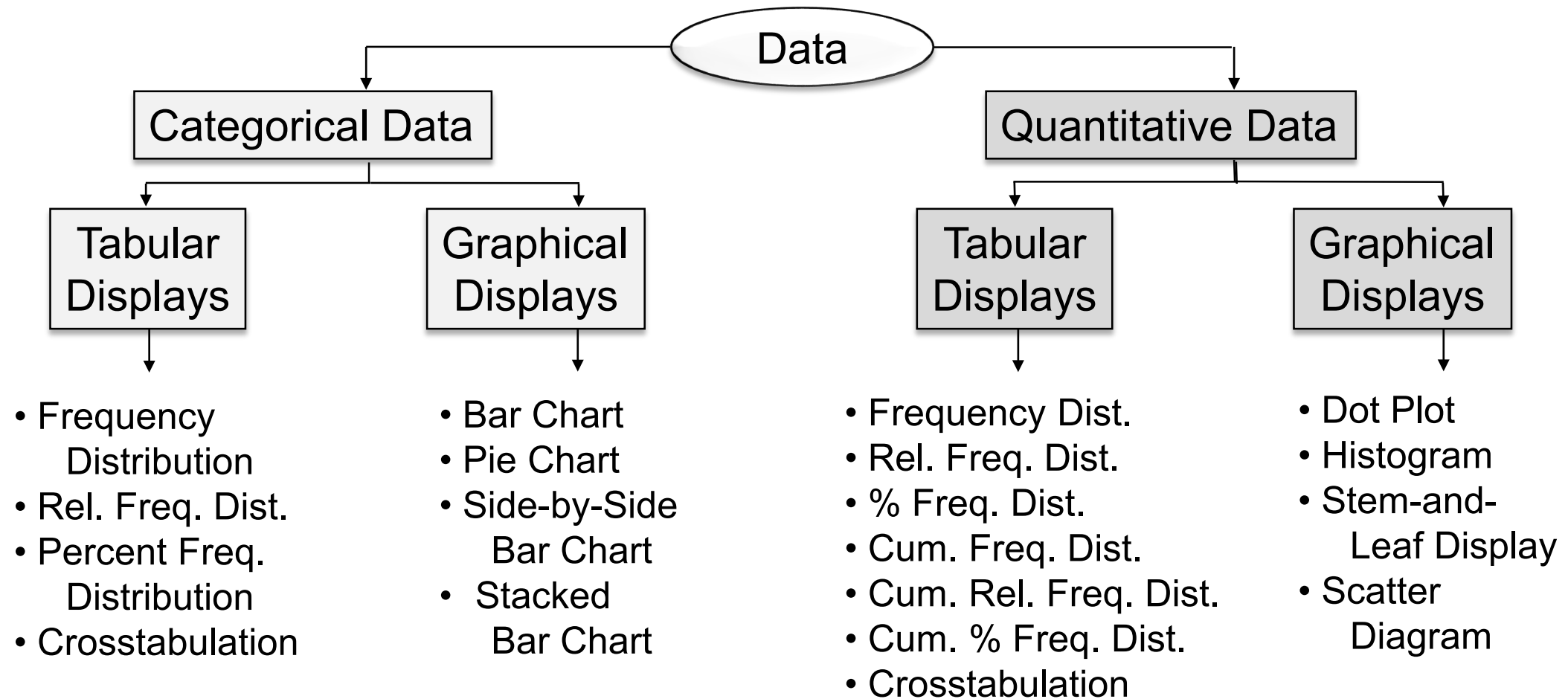
Data Dashboards

- A data dashboard is a widely used data visualization tool.
- It organizes and presents key performance indicators (KPIs) used to monitor an organization or process.
- It provides timely summary information that is easy to read, understand, and interpret.
- Some additional guidelines include . . .
 - Minimize the need for screen scrolling.
 - Avoid unnecessary use of color or 3D displays.
 - Use borders between charts to improve readability.

Data Dashboard Example



Tabular and Graphical Displays



End of Chapter 2, Part B

